US009240049B2

(12) **United States Patent**
Ciurea et al.

(10) **Patent No.: US 9,240,049 B2**
(45) **Date of Patent: *Jan. 19, 2016**

(54) **SYSTEMS AND METHODS FOR MEASURING DEPTH USING AN ARRAY OF INDEPENDENTLY CONTROLLABLE CAMERAS**

(71) Applicant: **Pelican Imaging Corporation**, Santa Clara, CA (US)

(72) Inventors: **Florian Ciurea**, San Jose, CA (US); **Kartik Venkataraman**, San Jose, CA (US); **Gabriel Molina**, Palo Alto, CA (US); **Dan Lelescu**, Morgan Hill, CA (US)

(73) Assignee: **Pelican Imaging Corporation**, Santa Clara, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **14/329,754**

(22) Filed: **Jul. 11, 2014**

(65) **Prior Publication Data**

US 2014/0321712 A1 Oct. 30, 2014

Related U.S. Application Data

(63) Continuation of application No. 14/144,458, filed on Dec. 30, 2013, now Pat. No. 8,780,113, which is a continuation of application No. 13/972,881, filed on Aug. 21, 2013, now Pat. No. 8,619,082.

(60) Provisional application No. 61/780,906, filed on Mar. 13, 2013, provisional application No. 61/691,666, filed on Aug. 21, 2012.

(51) **Int. Cl.**
G06T 7/00 (2006.01)
G02B 27/00 (2006.01)
(Continued)

(52) **U.S. Cl.**
CPC ***G06T 7/0075*** (2013.01); ***G02B 27/0075*** (2013.01); ***G06T 7/002*** (2013.01);
(Continued)

(58) **Field of Classification Search**
CPC G06T 2207/10012; G06T 2207/20228; G06T 7/0022; G06T 7/0065; G06T 2200/08; G06T 7/0075; G06T 3/4038; G06T 7/0077; H04N 13/0242; H04N 2013/0081; H04N 13/0037; H04N 2013/0088
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,124,798 A 11/1978 Thompson
4,198,646 A 4/1980 Alexander et al.
(Continued)

FOREIGN PATENT DOCUMENTS

EP 840502 A2 5/1998
EP 2336816 A2 6/2011
(Continued)

OTHER PUBLICATIONS

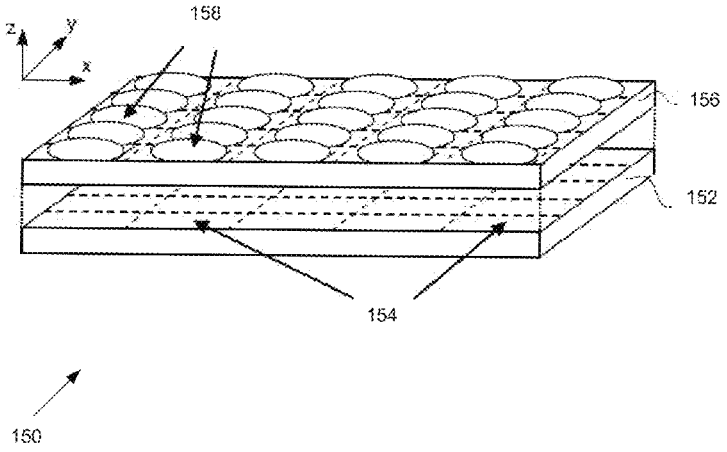
Bose et al., "Superresolution and Noise Filtering Using Moving Least Squares", IEEE Transactions on Image Processing, date unknown, 21 pgs, 2006.
(Continued)

Primary Examiner — Haixia Du
(74) *Attorney, Agent, or Firm* — KPPB LLP

(57) **ABSTRACT**

Systems in accordance with embodiments of the invention can perform parallax detection and correction in images captured using array cameras. Due to the different viewpoints of the cameras, parallax results in variations in the position of objects within the captured images of the scene. Methods in accordance with embodiments of the invention provide an accurate account of the pixel disparity due to parallax between the different cameras in the array, so that appropriate scene-dependent geometric shifts can be applied to the pixels of the captured images when performing super-resolution processing. In a number of embodiments, generating depth estimates considers the similarity of pixels in multiple spectral channels. In certain embodiments, generating depth estimates involves generating a confidence map indicating the reliability of depth estimates.

20 Claims, 54 Drawing Sheets



(51) **Int. Cl.**

G06T 15/20	(2011.01)
H04N 13/02	(2006.01)
H04N 9/097	(2006.01)
H04N 13/00	(2006.01)

(52) **U.S. Cl.**

CPC ***G06T 7/0065*** (2013.01); ***G06T 15/20*** (2013.01); ***H04N 9/097*** (2013.01); ***H04N 13/0022*** (2013.01); ***H04N 13/0232*** (2013.01); ***H04N 13/0242*** (2013.01); *G06T 2200/21* (2013.01); *G06T 2207/10012* (2013.01); *G06T 2207/10024* (2013.01); *G06T 2207/10052* (2013.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,323,925	A	4/1982	Abell et al.
4,460,449	A	7/1984	Montalbano
4,467,365	A	8/1984	Murayama et al.
5,005,083	A	4/1991	Grage et al.
5,070,414	A	12/1991	Tsutsumi
5,144,448	A	9/1992	Hornbaker
5,327,125	A	7/1994	Iwase et al.
5,629,524	A	5/1997	Stettner et al.
5,808,350	A	9/1998	Jack et al.
5,832,312	A	11/1998	Rieger et al.
5,880,691	A	3/1999	Fossum et al.
5,933,190	A	8/1999	Dierickx et al.
5,973,844	A	10/1999	Burger
6,002,743	A	12/1999	Telymonde
6,005,607	A	12/1999	Uomori et al.
6,034,690	A	3/2000	Gallery et al.
6,069,351	A	5/2000	Mack
6,069,365	A	5/2000	Chow et al.
6,097,394	A	8/2000	Levoy et al.
6,124,974	A	9/2000	Burger
6,137,535	A	10/2000	Meyers
6,141,048	A	10/2000	Meyers
6,160,909	A	12/2000	Melen
6,163,414	A	12/2000	Kikuchi et al.
6,172,352	B1	1/2001	Liu et al.
6,175,379	B1	1/2001	Uomori et al.
6,205,241	B1	3/2001	Melen
6,239,909	B1	5/2001	Hayashi et al.
6,358,862	B1	3/2002	Ireland et al.
6,443,579	B1	9/2002	Myers et al.
6,476,805	B1	11/2002	Shum et al.
6,477,260	B1	11/2002	Shimomura
6,525,302	B2	2/2003	Dowski, Jr. et al.
6,563,537	B1	5/2003	Kawamura et al.
6,571,466	B1	6/2003	Glenn et al.
6,603,513	B1	8/2003	Berezin
6,611,289	B1	8/2003	Yu
6,627,896	B1	9/2003	Hashimoto et al.
6,628,330	B1	9/2003	Lin
6,635,941	B2	10/2003	Suda
6,639,596	B1	10/2003	Shum et al.
6,657,218	B2	12/2003	Noda
6,671,399	B1	12/2003	Berestov
6,750,904	B1	6/2004	Lambert
6,765,617	B1	7/2004	Tangen et al.
6,771,833	B1	8/2004	Edgar
6,774,941	B1	8/2004	Boisvert et al.
6,795,253	B2	9/2004	Shinohara
6,819,358	B1	11/2004	Kagle et al.
6,879,735	B1	4/2005	Portniaguine et al.
6,903,770	B1	6/2005	Kobayashi et al.
6,909,121	B2	6/2005	Nishikawa
6,927,922	B2	8/2005	George et al.
6,958,862	B1	10/2005	Joseph
7,085,409	B2	8/2006	Sawhney et al.
7,161,614	B1	1/2007	Yamashita et al.
7,199,348	B2	4/2007	Olsen et al.
7,262,799	B2	8/2007	Suda
7,292,735	B2	11/2007	Blake et al.
7,295,697	B1	11/2007	Satoh
7,369,165	B2	5/2008	Bosco et al.
7,391,572	B2	6/2008	Jacobowitz et al.
7,408,725	B2	8/2008	Sato
7,425,984	B2	9/2008	Chen
7,606,484	B1	10/2009	Richards et al.
7,633,511	B2	12/2009	Shum et al.
7,639,435	B2	12/2009	Chiang et al.
7,646,549	B2	1/2010	Zalevsky et al.
7,657,090	B2	2/2010	Omatsu et al.
7,675,080	B2	3/2010	Boettiger
7,675,681	B2	3/2010	Tomikawa et al.
7,706,634	B2	4/2010	Schmitt et al.
7,723,662	B2	5/2010	Levoy et al.
7,782,364	B2	8/2010	Smith
7,840,067	B2	11/2010	Shen et al.
7,912,673	B2	3/2011	Hébert et al.
7,986,018	B2	7/2011	Rennie
7,990,447	B2	8/2011	Honda et al.
8,000,498	B2	8/2011	Shih et al.
8,013,904	B2	9/2011	Tan et al.
8,027,531	B2	9/2011	Wilburn et al.
8,044,994	B2	10/2011	Vetro et al.
8,077,245	B2	12/2011	Adamo et al.
8,098,297	B2	1/2012	Crisan et al.
8,098,304	B2	1/2012	Pinto et al.
8,106,949	B2	1/2012	Tan et al.
8,126,279	B2	2/2012	Marcellin et al.
8,130,120	B2	3/2012	Kawabata et al.
8,131,097	B2	3/2012	Lelescu et al.
8,164,629	B1	4/2012	Zhang
8,169,486	B2	5/2012	Corcoran et al.
8,180,145	B2	5/2012	Wu et al.
8,189,065	B2	5/2012	Georgiev et al.
8,189,089	B1	5/2012	Georgiev et al.
8,212,914	B2	7/2012	Chiu
8,213,711	B2	7/2012	Tam et al.
8,231,814	B2	7/2012	Duparre
8,242,426	B2	8/2012	Ward et al.
8,244,027	B2	8/2012	Takahashi
8,244,058	B1	8/2012	Intwala et al.
8,254,668	B2	8/2012	Mashitani et al.
8,279,325	B2	10/2012	Pitts et al.
8,280,194	B2	10/2012	Wong et al.
8,289,409	B2	10/2012	Chang
8,289,440	B2	10/2012	Pitts et al.
8,290,358	B1	10/2012	Georgiev
8,294,099	B2	10/2012	Blackwell, Jr.
8,305,456	B1	11/2012	McMahon
8,315,476	B1	11/2012	Georgiev et al.
8,345,144	B1	1/2013	Georgiev et al.
8,360,574	B2	1/2013	Ishak et al.
8,406,562	B2	3/2013	Bassi et al.
8,446,492	B2	5/2013	Nakano et al.
8,456,517	B2	6/2013	Mor et al.
8,493,496	B2	7/2013	Freedman et al.
8,514,491	B2	8/2013	Duparre
8,541,730	B2	9/2013	Inuiya
8,542,933	B2	9/2013	Venkataraman et al.
8,553,093	B2	10/2013	Wong et al.
8,559,756	B2	10/2013	Georgiev et al.
8,581,995	B2	11/2013	Lin et al.
8,619,082	B1	12/2013	Ciurea et al.
8,648,918	B2	2/2014	Kauker et al.
8,655,052	B2	2/2014	Spooner et al.
8,682,107	B2	3/2014	Yoon et al.
8,692,893	B2	4/2014	McMahon
8,773,536	B1	7/2014	Zhang
8,780,113	B1	7/2014	Ciurea et al.
8,804,255	B2	8/2014	Duparre
8,830,375	B2	9/2014	Ludwig
8,831,367	B2	9/2014	Venkataraman et al.
8,842,201	B2	9/2014	Tajiri
8,854,462	B2	10/2014	Herbin et al.
8,861,089	B2	10/2014	Duparre
8,866,912	B2	10/2014	Mullis
8,866,920	B2	10/2014	Venkataraman et al.
8,866,951	B2	10/2014	Keelan

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,878,950 B2 11/2014 Lelescu et al.
8,885,059 B1 11/2014 Venkataraman et al.
8,896,594 B2 11/2014 Xiong et al.
8,896,719 B1 11/2014 Venkataraman et al.
8,902,321 B2 12/2014 Venkataraman et al.
9,019,426 B2 4/2015 Han et al.
9,030,528 B2 5/2015 Shpunt et al.
2001/0005225 A1 6/2001 Clark et al.
2001/0019621 A1 9/2001 Hanna et al.
2001/0038387 A1 11/2001 Tomooka et al.
2002/0012056 A1 1/2002 Trevino et al.
2002/0027608 A1 3/2002 Johnson et al.
2002/0039438 A1 4/2002 Mori et al.
2002/0063807 A1 5/2002 Margulis
2002/0087403 A1 7/2002 Meyers et al.
2002/0089596 A1 7/2002 Suda
2002/0094027 A1 7/2002 Sato et al.
2002/0101528 A1 8/2002 Lee et al.
2002/0113867 A1 8/2002 Takigawa et al.
2002/0113888 A1 8/2002 Sonoda et al.
2002/0163054 A1 11/2002 Suda
2002/0167537 A1 11/2002 Trajkovic
2002/0177054 A1 11/2002 Saitoh et al.
2003/0086079 A1 5/2003 Barth et al.
2003/0124763 A1 7/2003 Fan et al.
2003/0140347 A1 7/2003 Varsa
2003/0179418 A1 9/2003 Wengender et al.
2003/0190072 A1 10/2003 Adkins et al.
2003/0211405 A1 11/2003 Venkataraman
2004/0008271 A1 1/2004 Hagimori et al.
2004/0012689 A1 1/2004 Tinnerino et al.
2004/0027358 A1 2/2004 Nakao
2004/0047274 A1 3/2004 Amanai
2004/0050104 A1 3/2004 Ghosh et al.
2004/0056966 A1 3/2004 Schechner et al.
2004/0066454 A1 4/2004 Otani et al.
2004/0096119 A1 5/2004 Williams
2004/0100570 A1 5/2004 Shizukuishi
2004/0114807 A1 6/2004 Lelescu et al.
2004/0151401 A1 8/2004 Sawhney et al.
2004/0165090 A1 8/2004 Ning
2004/0169617 A1 9/2004 Yelton et al.
2004/0170340 A1 9/2004 Tipping et al.
2004/0174439 A1 9/2004 Upton
2004/0207836 A1 10/2004 Chhibber et al.
2004/0213449 A1 10/2004 Safaee-Rad et al.
2004/0218809 A1 11/2004 Blake et al.
2004/0234873 A1 11/2004 Venkataraman
2004/0240052 A1 12/2004 Minefuji et al.
2004/0251509 A1 12/2004 Choi
2004/0264806 A1 12/2004 Herley
2005/0006477 A1 1/2005 Patel
2005/0009313 A1 1/2005 Suzuki et al.
2005/0012035 A1 1/2005 Miller
2005/0036778 A1 2/2005 DeMonte
2005/0047678 A1 3/2005 Jones et al.
2005/0048690 A1 3/2005 Yamamoto
2005/0068436 A1 3/2005 Fraenkel et al.
2005/0132098 A1 6/2005 Sonoda et al.
2005/0134712 A1 6/2005 Gruhlke et al.
2005/0147277 A1 7/2005 Higaki et al.
2005/0151759 A1 7/2005 Gonzalez-Banos et al.
2005/0175257 A1 8/2005 Kuroki
2005/0185711 A1 8/2005 Pfister et al.
2005/0205785 A1 9/2005 Hornback et al.
2005/0219363 A1 10/2005 Kohler et al.
2005/0225654 A1 10/2005 Feldman et al.
2005/0275946 A1 12/2005 Choo et al.
2005/0286612 A1 12/2005 Takanashi
2006/0002635 A1 1/2006 Nestares et al.
2006/0023197 A1 2/2006 Joel
2006/0023314 A1 2/2006 Boettiger et al.
2006/0033005 A1 2/2006 Jerdev et al.
2006/0034003 A1 2/2006 Zalevsky
2006/0038891 A1 2/2006 Okutomi et al.
2006/0039611 A1 2/2006 Rother
2006/0049930 A1 3/2006 Zruya et al.
2006/0054780 A1 3/2006 Garrood et al.
2006/0054782 A1 3/2006 Olsen et al.
2006/0055811 A1 3/2006 Frtiz et al.
2006/0069478 A1 3/2006 Iwama
2006/0072029 A1 4/2006 Miyatake et al.
2006/0087747 A1 4/2006 Ohzawa et al.
2006/0098888 A1 5/2006 Morishita
2006/0125936 A1 6/2006 Gruhike et al.
2006/0138322 A1 6/2006 Costello et al.
2006/0152803 A1 7/2006 Provitola
2006/0157640 A1 7/2006 Perlman et al.
2006/0159369 A1 7/2006 Young
2006/0176566 A1 8/2006 Boettiger et al.
2006/0187338 A1 8/2006 May et al.
2006/0197937 A1 9/2006 Bamji et al.
2006/0203113 A1 9/2006 Wada et al.
2006/0210186 A1 9/2006 Berkner
2006/0239549 A1 10/2006 Kelly et al.
2006/0243889 A1 11/2006 Farnworth et al.
2006/0251410 A1 11/2006 Trutna
2006/0274174 A1 12/2006 Tewinkle
2006/0278948 A1 12/2006 Yamaguchi et al.
2006/0279648 A1 12/2006 Senba et al.
2007/0002159 A1 1/2007 Olsen et al.
2007/0024614 A1 2/2007 Tam
2007/0036427 A1 2/2007 Nakamura et al.
2007/0040828 A1 2/2007 Zalevsky et al.
2007/0040922 A1 2/2007 McKee et al.
2007/0041391 A1 2/2007 Lin et al.
2007/0052825 A1 3/2007 Cho
2007/0083114 A1 4/2007 Yang et al.
2007/0085917 A1 4/2007 Kobayashi
2007/0102622 A1 5/2007 Olsen et al.
2007/0126898 A1 6/2007 Feldman et al.
2007/0127831 A1 6/2007 Venkataraman
2007/0139333 A1 6/2007 Sato et al.
2007/0146511 A1 6/2007 Kinoshita et al.
2007/0158427 A1 7/2007 Zhu et al.
2007/0159541 A1 7/2007 Sparks et al.
2007/0160310 A1 7/2007 Tanida et al.
2007/0165931 A1 7/2007 Higaki
2007/0171290 A1 7/2007 Kroger
2007/0206241 A1 9/2007 Smith et al.
2007/0211164 A1 9/2007 Olsen et al.
2007/0216765 A1 9/2007 Wong et al.
2007/0228256 A1 10/2007 Mentzer et al.
2007/0257184 A1 11/2007 Olsen et al.
2007/0258006 A1 11/2007 Olsen et al.
2007/0258706 A1 11/2007 Raskar et al.
2007/0263114 A1 11/2007 Gurevich et al.
2007/0268374 A1 11/2007 Robinson
2007/0296832 A1 12/2007 Ota et al.
2007/0296835 A1 12/2007 Olsen et al.
2007/0296847 A1 12/2007 Chang et al.
2008/0019611 A1 1/2008 Larkin et al.
2008/0025649 A1 1/2008 Liu et al.
2008/0030597 A1 2/2008 Olsen et al.
2008/0043095 A1 2/2008 Vetro et al.
2008/0043096 A1 2/2008 Vetro et al.
2008/0062164 A1 3/2008 Bassi et al.
2008/0079805 A1 4/2008 Takagi et al.
2008/0080028 A1 4/2008 Bakin et al.
2008/0084486 A1 4/2008 Enge et al.
2008/0088793 A1 4/2008 Sverdrup et al.
2008/0112635 A1 5/2008 Kondo et al.
2008/0118241 A1 5/2008 Tekolste et al.
2008/0131019 A1 6/2008 Ng
2008/0131107 A1 6/2008 Ueno
2008/0151097 A1 6/2008 Chen et al.
2008/0152215 A1 6/2008 Horie et al.
2008/0152296 A1 6/2008 Oh et al.
2008/0158259 A1 7/2008 Kempf et al.
2008/0158375 A1 7/2008 Kakkori et al.
2008/0158698 A1 7/2008 Chang et al.
2008/0187305 A1 8/2008 Raskar et al.
2008/0193026 A1 8/2008 Horie et al.
2008/0218610 A1 9/2008 Chapman et al.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2008/0219654 A1 9/2008 Border et al.
2008/0239116 A1 10/2008 Smith
2008/0240598 A1 10/2008 Hasegawa
2008/0247638 A1 10/2008 Tanida et al.
2008/0247653 A1 10/2008 Moussavi et al.
2008/0272416 A1 11/2008 Yun
2008/0273751 A1 11/2008 Yuan et al.
2008/0278591 A1 11/2008 Barna et al.
2008/0298674 A1 12/2008 Baker et al.
2009/0050946 A1 2/2009 Duparre et al.
2009/0052743 A1 2/2009 Techmer
2009/0060281 A1 3/2009 Tanida et al.
2009/0086074 A1 4/2009 Li et al.
2009/0091806 A1 4/2009 Inuiya
2009/0096050 A1 4/2009 Park
2009/0102956 A1 4/2009 Georgiev
2009/0109306 A1 4/2009 Shan et al.
2009/0128833 A1 5/2009 Yahav
2009/0140131 A1 6/2009 Utagawa
2009/0152664 A1 6/2009 Klem et al.
2009/0167922 A1 7/2009 Perlman et al.
2009/0179142 A1 7/2009 Duparre et al.
2009/0180021 A1 7/2009 Kikuchi et al.
2009/0200622 A1 8/2009 Tai et al.
2009/0201371 A1 8/2009 Matsuda et al.
2009/0207235 A1 8/2009 Francini et al.
2009/0225203 A1 9/2009 Tanida et al.
2009/0237520 A1 9/2009 Kaneko et al.
2009/0245573 A1 10/2009 Saptharishi et al.
2009/0263017 A1 10/2009 Tanbakuchi
2009/0268192 A1 10/2009 Koenck et al.
2009/0268970 A1 10/2009 Babacan et al.
2009/0268983 A1 10/2009 Stone
2009/0274387 A1 11/2009 Jin
2009/0284651 A1 11/2009 Srinivasan
2009/0297056 A1 12/2009 Lelescu et al.
2009/0302205 A9 12/2009 Olsen et al.
2009/0323195 A1 12/2009 Hembree et al.
2009/0323206 A1 12/2009 Oliver et al.
2009/0324118 A1 12/2009 Maslov et al.
2010/0002126 A1 1/2010 Wenstrand et al.
2010/0002313 A1 1/2010 Duparre et al.
2010/0002314 A1 1/2010 Duparre
2010/0013927 A1 1/2010 Nixon
2010/0053342 A1 3/2010 Hwang et al.
2010/0053600 A1 3/2010 Tanida et al.
2010/0060746 A9 3/2010 Olsen et al.
2010/0074532 A1 3/2010 Gordon et al.
2010/0085425 A1 4/2010 Tan
2010/0086227 A1 4/2010 Sun et al.
2010/0091389 A1 4/2010 Henriksen et al.
2010/0097491 A1 4/2010 Farina et al.
2010/0103259 A1 4/2010 Tanida et al.
2010/0103308 A1 4/2010 Butterfield et al.
2010/0111444 A1 5/2010 Coffman
2010/0118127 A1 5/2010 Nam et al.
2010/0128145 A1 5/2010 Pitts et al.
2010/0133230 A1 6/2010 Henriksen et al.
2010/0133418 A1 6/2010 Sargent et al.
2010/0141802 A1 6/2010 Knight et al.
2010/0142839 A1 6/2010 Lakus-Becker
2010/0157073 A1 6/2010 Kondo et al.
2010/0165152 A1 7/2010 Lim
2010/0166410 A1 7/2010 Chang et al.
2010/0177411 A1 7/2010 Hegde et al.
2010/0194901 A1 8/2010 van Hoorebeke et al.
2010/0195716 A1 8/2010 Klein et al.
2010/0201834 A1 8/2010 Maruyama et al.
2010/0208100 A9 8/2010 Olsen et al.
2010/0220212 A1 9/2010 Perlman et al.
2010/0231285 A1 9/2010 Boomer et al.
2010/0244165 A1 9/2010 Lake et al.
2010/0265385 A1 10/2010 Knight et al.
2010/0281070 A1 11/2010 Chan et al.
2010/0302423 A1 12/2010 Adams, Jr. et al.
2010/0309292 A1 12/2010 Ho et al.
2011/0001037 A1 1/2011 Tewinkle
2011/0018973 A1 1/2011 Takayama
2011/0019243 A1 1/2011 Constant, Jr. et al.
2011/0032370 A1 2/2011 Ludwig
2011/0043661 A1 2/2011 Podoleanu
2011/0043665 A1 2/2011 Ogasahara
2011/0043668 A1 2/2011 McKinnon et al.
2011/0069189 A1 3/2011 Venkataraman et al.
2011/0080487 A1 4/2011 Venkataraman et al.
2011/0108708 A1 5/2011 Olsen et al.
2011/0121421 A1 5/2011 Charbon et al.
2011/0122308 A1 5/2011 Duparre
2011/0128393 A1 6/2011 Tavi et al.
2011/0128412 A1 6/2011 Milnes et al.
2011/0149408 A1 6/2011 Hahgholt et al.
2011/0149409 A1 6/2011 Haugholt et al.
2011/0153248 A1 6/2011 Gu et al.
2011/0157321 A1 6/2011 Nakajima et al.
2011/0176020 A1 7/2011 Chang
2011/0211824 A1 9/2011 Georgiev et al.
2011/0221599 A1 9/2011 Högasten
2011/0221658 A1 9/2011 Haddick et al.
2011/0221939 A1 9/2011 Jerdev
2011/0234841 A1 9/2011 Akeley et al.
2011/0241234 A1 10/2011 Duparre
2011/0242342 A1 10/2011 Goma et al.
2011/0242355 A1 10/2011 Goma et al.
2011/0242356 A1 10/2011 Aleksic et al.
2011/0255592 A1 10/2011 Sung et al.
2011/0255745 A1 10/2011 Hodder et al.
2011/0261993 A1 10/2011 Weiming et al.
2011/0267348 A1 11/2011 Lin et al.
2011/0273531 A1 11/2011 Ito et al.
2011/0274366 A1 11/2011 Tardif
2011/0279721 A1 11/2011 Mcmahon
2011/0285866 A1 11/2011 Bhrugumalla et al.
2011/0285910 A1 11/2011 Bamji et al.
2011/0298917 A1 12/2011 Yanagita
2011/0300929 A1 12/2011 Tardif et al.
2011/0310980 A1 12/2011 Mathew
2011/0316968 A1 12/2011 Taguchi et al.
2011/0317766 A1 12/2011 Lim, II et al.
2012/0012748 A1 1/2012 Pain et al.
2012/0023456 A1 1/2012 Sun et al.
2012/0026297 A1 2/2012 Sato
2012/0026342 A1 2/2012 Yu et al.
2012/0039525 A1 2/2012 Tian et al.
2012/0044249 A1 2/2012 Mashitani et al.
2012/0044372 A1 2/2012 Côté et al.
2012/0069235 A1 3/2012 Imai
2012/0105691 A1 5/2012 Waqas et al.
2012/0113413 A1 5/2012 Miahczylowicz-Wolski et al.
2012/0147139 A1 6/2012 Li et al.
2012/0147205 A1 6/2012 Lelescu et al.
2012/0153153 A1 6/2012 Chang et al.
2012/0154551 A1 6/2012 Inoue
2012/0170134 A1 7/2012 Bolis et al.
2012/0176479 A1 7/2012 Mayhew et al.
2012/0188389 A1 7/2012 Lin et al.
2012/0188420 A1 7/2012 Black et al.
2012/0188634 A1 7/2012 Kubala et al.
2012/0198677 A1 8/2012 Duparre
2012/0200734 A1 8/2012 Tang
2012/0219236 A1 8/2012 Ali et al.
2012/0229628 A1 9/2012 Ishiyama et al.
2012/0249550 A1 10/2012 Akeley et al.
2012/0249836 A1 10/2012 Ali et al.
2012/0262607 A1 10/2012 Shimura et al.
2012/0268574 A1 10/2012 Gidon et al.
2012/0287291 A1 11/2012 McMahon et al.
2012/0293695 A1 11/2012 Tanaka
2012/0314033 A1 12/2012 Lee et al.
2012/0327222 A1 12/2012 Ng et al.
2013/0002828 A1 1/2013 Ding et al.
2013/0003184 A1 1/2013 Duparre
2013/0010073 A1 1/2013 Do et al.
2013/0016885 A1 1/2013 Tsujimoto et al.
2013/0022111 A1 1/2013 Chen et al.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2013/0027580	A1	1/2013	Olsen et al.
2013/0033579	A1	2/2013	Wajs
2013/0050504	A1	2/2013	Safaee-Rad et al.
2013/0050526	A1	2/2013	Keelan
2013/0057710	A1	3/2013	McMahon
2013/0070060	A1	3/2013	Chatterjee et al.
2013/0077880	A1	3/2013	Venkataraman et al.
2013/0077882	A1	3/2013	Venkataraman et al.
2013/0088489	A1	4/2013	Schmeitz et al.
2013/0088637	A1	4/2013	Duparre
2013/0113899	A1	5/2013	Morohoshi et al.
2013/0120605	A1	5/2013	Georgiev et al.
2013/0128068	A1	5/2013	Georgiev et al.
2013/0128069	A1	5/2013	Georgiev et al.
2013/0128087	A1	5/2013	Georgiev et al.
2013/0128121	A1	5/2013	Agarwala et al.
2013/0147979	A1	6/2013	McMahon et al.
2013/0215108	A1	8/2013	McMahon et al.
2013/0222556	A1	8/2013	Shimada
2013/0223759	A1	8/2013	Nishiyama et al.
2013/0229540	A1	9/2013	Farina et al.
2013/0230237	A1	9/2013	Schlosser, Markus et al.
2013/0259317	A1	10/2013	Gaddy
2013/0265459	A1	10/2013	Duparre et al.
2013/0274923	A1	10/2013	By et al.
2013/0293760	A1	11/2013	Nisenzon et al.
2014/0009586	A1	1/2014	McNamer et al.
2014/0076336	A1	3/2014	Clayton et al.
2014/0079336	A1	3/2014	Venkataraman et al.
2014/0092281	A1	4/2014	Nisenzon et al.
2014/0104490	A1	4/2014	Hsieh et al.
2014/0118493	A1	5/2014	Sali et al.
2014/0132810	A1	5/2014	McMahon
2014/0176592	A1	6/2014	Wilburn et al.
2014/0192253	A1	7/2014	Laroia
2014/0198188	A1	7/2014	Izawa
2014/0218546	A1	8/2014	Mcmahon
2014/0232822	A1	8/2014	Venkataraman et al.
2014/0240528	A1	8/2014	Venkataraman et al.
2014/0240529	A1	8/2014	Venkataraman et al.
2014/0253738	A1	9/2014	Mullis
2014/0267243	A1	9/2014	Venkataraman et al.
2014/0267286	A1	9/2014	Duparre
2014/0267633	A1	9/2014	Venkataraman et al.
2014/0267762	A1	9/2014	Mullis et al.
2014/0267890	A1	9/2014	Lelescu et al.
2014/0285675	A1	9/2014	Mullis
2014/0313315	A1	10/2014	Shoham et al.
2014/0321712	A1	10/2014	Ciurea et al.
2014/0333731	A1	11/2014	Venkataraman et al.
2014/0333764	A1	11/2014	Venkataraman et al.
2014/0333787	A1	11/2014	Venkataraman et al.
2014/0340539	A1	11/2014	Venkataraman et al.
2014/0347509	A1	11/2014	Venkataraman et al.
2014/0368686	A1	12/2014	Duparre
2015/0035992	A1	2/2015	Mullis
2015/0042766	A1	2/2015	Ciurea et al.
2015/0042767	A1	2/2015	Ciurea et al.
2015/0049915	A1	2/2015	Ciurea et al.
2015/0049916	A1	2/2015	Ciurea et al.
2015/0049917	A1	2/2015	Ciurea et al.

FOREIGN PATENT DOCUMENTS

JP	2006033493	A	2/2006
JP	2007520107	A	7/2007
JP	2011109484	A	6/2011
JP	2013526801	A	6/2013
JP	2014521117	A	8/2014
KR	1020110097647	A	8/2011
TW	200939739	A	9/2009
WO	2007083579	A1	7/2007
WO	2008108271	A1	9/2008
WO	2009151903	A2	12/2009
WO	2011063347	A2	5/2011
WO	2011116203	A1	9/2011
WO	2011063347	A3	10/2011
WO	2011143501	A1	11/2011
WO	2012057619	A1	5/2012
WO	2012057620	A2	5/2012
WO	2012057621	A1	5/2012
WO	2012057622	A1	5/2012
WO	2012057623	A1	5/2012
WO	2012057620	A3	6/2012
WO	2012074361	A1	6/2012
WO	2012078126	A1	6/2012
WO	2012082904	A1	6/2012
WO	2012155119	A1	11/2012
WO	2013003276	A1	1/2013
WO	2013043751	A1	3/2013
WO	2013043761	A1	3/2013
WO	2013049699	A1	4/2013
WO	2013055960	A1	4/2013
WO	2013119706	A1	8/2013
WO	2013126578	A1	8/2013
WO	2014052974	A2	4/2014
WO	2014032020	A3	5/2014
WO	2014078443	A1	5/2014
WO	2014130849	A1	8/2014
WO	2014133974	A1	9/2014
WO	2014138695	A1	9/2014
WO	2014138697	A1	9/2014
WO	2014144157	A1	9/2014
WO	2014145856	A1	9/2014
WO	2014149403	A1	9/2014
WO	2014150856	A1	9/2014
WO	2014159721	A1	10/2014
WO	2014159779	A1	10/2014
WO	2014160142	A1	10/2014
WO	2014164550	A2	10/2014
WO	2014164909	A1	10/2014
WO	2014165244	A1	10/2014

OTHER PUBLICATIONS

Boye et al., "Comparison of Subpixel Image Registration Algorithms", Proc. of SPIE—IS&T Electronic Imaging, vol. 7246, pp. 72460X-1-72460X-9, 2009.

Bruckner et al., "Artificial compound eye applying hyperacuity", Optics Express, Dec. 11, 2006, vol. 14, No. 25, pp. 12076-12084.

Bruckner et al., "Driving microoptical imaging systems towards miniature camera applications", Proc. SPIE, Micro-Optics, 2010, 11 pgs.

Bruckner et al., "Thin wafer-level camera lenses inspired by insect compound eyes", Optics Express, Nov. 22, 2010, vol. 18, No. 24, pp. 24379-24394.

Capel, "Image Mosaicing and Super-resolution", [online], Retrieved on Nov. 10, 2012. Retrieved from the Internet at URL:<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.226.2643&rep=rep1 &type=pdf>, Title pg., abstract, table of contents, pp. 1-263 (269 total pages).

Chan et al., "Extending the Depth of Field in a Compound-Eye Imaging System with Super-Resolution Reconstruction", Proceedings—International Conference on Pattern Recognition, 2006, vol. 3, pp. 623-626.

Chan et al., "Investigation of Computational Compound-Eye Imaging System with Super-Resolution Reconstruction", IEEE, ISASSP 2006, pp. 1177-1180.

Chan et al., "Super-resolution reconstruction in a computational compound-eye imaging system", Multidim Syst Sign Process, 2007, vol. 18, pp. 83-101.

Chen et al., "Interactive deformation of light fields", In Proceedings of SIGGRAPH I3D 2005, pp. 139-146.

Drouin et al., "Fast Multiple-Baseline Stereo with Occlusion", Proceedings of the Fifth International Conference on 3-D Digital Imaging and Modeling, 2005, 8 pgs.

Drouin et al., "Geo-Consistency for Wide Multi-Camera Stereo", Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, 8 pgs.

Drouin et al., "Improving Border Localization of Multi-Baseline Stereo Using Border-Cut", International Journal of Computer Vision, Jul. 2009, vol. 83, Issue 3, 8 pgs.

(56) **References Cited**

OTHER PUBLICATIONS

Duparre et al., "Artificial apposition compound eye fabricated by micro-optics technology", Applied Optics, Aug. 1, 2004, vol. 43, No. 22, pp. 4303-4310.

Duparre et al., "Artificial compound eye zoom camera", Bioinspiration & Biomimetics, 2008, vol. 3, pp. 1-6.

Duparre et al., "Artificial compound eyes—different concepts and their application to ultra flat image acquisition sensors", MOEMS and Miniaturized Systems IV, Proc. SPIE 5346, Jan. 2004, pp. 89-100.

Duparre et al., "Chirped arrays of refractive ellipsoidal microlenses for aberration correction under oblique incidence", Optics Express, Dec. 26, 2005, vol. 13, No. 26, pp. 10539-10551.

Duparre et al., "Micro-optical artificial compound eyes", Bioinspiration & Biomimetics, 2006, vol. 1, pp. R1-R16.

Duparre et al., "Microoptical artificial compound eyes—from design to experimental verification of two different concepts", Proc. of SPIE, Optical Design and Engineering II, vol. 5962, pp. 59622A-1-59622A-12, 2005.

Duparre et al., "Microoptical Artificial Compound Eyes—Two Different Concepts for Compact Imaging Systems", 11th Microoptics Conference, Oct. 30-Nov. 2, 2005, 2 pgs.

Duparre et al., "Microoptical telescope compound eye", Optics Express, Feb. 7, 2005, vol. 13, No. 3, pp. 889-903.

Duparre et al., "Micro-optically fabricated artificial apposition compound eye", Electronic Imaging—Science and Technology, Prod. SPIE 5301, Jan. 2004, pp. 25-33.

Duparre et al., "Novel Optics/Micro-Optics for Miniature Imaging Systems", Proc. of SPIE, 2006, vol. 6196, pp. 619607-1-619607-15.

Duparre et al., "Theoretical analysis of an artificial superposition compound eye for application in ultra flat digital image acquisition devices", Optical Systems Design, Proc. SPIE 5249, Sep. 2003, pp. 408-418.

Duparre et al., "Thin compound-eye camera", Applied Optics, May 20, 2005, vol. 44, No. 15, pp. 2949-2956.

Duparre et al., "Ultra-Thin Camera Based on Artificial Apposistion Compound Eyes", 10th Microoptics Conference, Sep. 1-3, 2004, 2 pgs.

Fanaswala, "Regularized Super-Resolution of Multi-View Images", Retrieved on Nov. 10, 2012. Retrieved from the Internet at URL:<http://www.site.uottawa.ca/-edubois/theses/Fanaswala_thesis.pdf>, 163 pgs.

Farrell et al., "Resolution and Light Sensitivity Tradeoff with Pixel Size", Proceedings of the SPIE Electronic Imaging 2006 Conference, 2006, vol. 6069, 8 pgs.

Farsiu et al., "Advances and Challenges in Super-Resolution", International Journal of Imaging Systems and Technology, 2004, vol. 14, pp. 47-57.

Farsiu et al., "Fast and Robust Multiframe Super Resolution", IEEE Transactions on Image Processing, Oct. 2004, vol. 13, No. 10, pp. 1327-1344.

Farsiu et al., "Multiframe Demosaicing and Super-Resolution of Color Images", IEEE Transactions on Image Processing, Jan. 2006, vol. 15, No. 1, pp. 141-159.

Feris et al., "Multi-Flash Stereopsis: Depth Edge Preserving Stereo with Small Baseline Illumination", IEEE Trans on PAMI, 2006, 31 pgs.

Fife et al., "A 3D Multi-Aperture Image Sensor Architecture", Custom Integrated Circuits Conference, 2006, CICC '06, IEEE, pp. 281-284.

Fife et al., "A 3MPixel Multi-Aperture Image Sensor with 0.7Mu Pixels in 0.11Mu CMOS", ISSCC 2008, Session 2, Image Sensors & Technology, 2008, pp. 48-50.

Fischer et al., Optical System Design, 2nd Edition, SPIE Press, pp. 191-198, 2008.

Fischer et al., "Optical System Design, 2nd Edition, SPIE Press, pp. 49-58", 2008.

Goldman et al., "Video Object Annotation, Navigation, and Composition", In Proceedings of UIST 2008, pp. 3-12.

Gortler et al., "The Lumigraph", In Proceedings of SIGGRAPH 1996, pp. 43-54.

Hacohen et al., "Non-Rigid Dense Correspondence with Applications for Image Enhancement", ACM Transactions on Graphics, 30, 4, 2011, pp. 70:1-70:10.

Hamilton, "JPEG File Interchange Format, Version 1.02", Sep. 1, 1992, 9 pgs.

Hardie, "A Fast Image Super-Algorithm Using an Adaptive Wiener Filter", IEEE Transactions on Image Processing, Dec. 2007, vol. 16, No. 12, pp. 2953-2964.

Hasinoff et al., "Search-and-Replace Editing for Personal Photo Collections", Computational Photography (ICCP) 2010, pp. 1-8.

Horisaki et al., "Irregular Lens Arrangement Design to Improve Imaging Performance of Compound-Eye Imaging Systems", Applied Physics Express, 2010, vol. 3, pp. 022501-1-022501-3.

Horisaki et al., "Superposition Imaging for Three-Dimensionally Space-Invariant Point Spread Functions", Applied Physics Express, 2011, vol. 4, pp. 112501-1-112501-3.

Horn et al., "LightShop: Interactive Light Field Manipulation and Rendering", In Proceedings of I3D 2007, pp. 121-128.

Isaksen et al., "Dynamically Reparameterized Light Fields", In Proceedings of SIGGRAPH 2000, pp. 297-306.

Jarabo et al., "Efficient Propagation of Light Field Edits", In Proceedings of SIACG 2011, pp. 75-80.

Joshi et al., "Synthetic Aperture Tracking: Tracking Through Occlusions", I CCV IEEE 11th International Conference on Computer Vision; Publication [online]. Oct. 2007 [retrieved Jul. 28, 2014]. Retrieved from the Internet: <URL: http:l/ieeexplore.ieee.org/stamp.jsp?tp=&arnumber=4409032&isnumber=4408819>; pp. 1-8.

Kang et al., "Handling Occlusions inn Dense Multi-View Stereo", Computer Vision and Pattern Recognition, 2001, vol. 1, pp. I-103-I-110.

Kitamura et al., "Reconstruction of a high-resolution image on a compound-eye image-capturing system", Applied Optics, Mar. 10, 2004, vol. 43, No. 8, pp. 1719-1727.

Chen et al., "KNN Matting", IEEE Transactions on Pattern Analysis and Machine Intelligence, Sep. 2013, vol. 35, No. 9, pp. 2175-2188.

Lai et al., "A Large-Scale Hierarchical Multi-View RGB-D Object Dataset", source and date unknown, 8 pgs, 2011.

Levin et al., "A Closed Form Solution to Natural Image Matting", Pattern Analysis and Machine Intelligence, Feb. 2008, vol. 30, 8 pgs.

Mitra et al., "Light Field Denoising, Light Field Superresolution and Stereo Camera Based Refocussing using a GMM Light Field Patch Prior", Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on Jun. 16-21, 2012, pp. 22-28.

Perwass et al., "Single Lens 3D-Camera with Extended Depth-of-Field", printed from www.raytrix.de, 15 pgs, 2012.

Tallon et al.,"Upsampling and Denoising of Depth Maps Via Joint-Segmentation", 20th European Signal Processing Conference, Aug. 27-31, 2012, 5 pgs.

Zhang, Qiang et al., "Depth estimation, spatially variant image registration, and super-resolution using a multi-lenslet camera", Proceedings of SPIE, vol. 7705, Apr. 23, 2010, pp. 770505-770505-8, XP055113797 ISSN: 0277-786X, DOI: 10.1117/12.852171.

International Preliminary Report on Patentability for International Application PCT/US2013/024987, Mailed Aug. 21, 2014, 13 Pgs.

International Preliminary Report on Patentability for International Application PCT/US2013/027146,Report Completed Apr. 2, 2013, 10 Pages.

International Search Report and Written Opinion for International Application No. PCT/US13/46002, Search Completed Nov. 13, 2013, 7 pgs.

International Search Report and Written Opinion for International Application No. PCT/US13/48772, Search Completed Oct. 21, 2013, 6 pgs.

International Search Report and Written Opinion for International Application No. PCT/US13/56065, Search Completed Nov. 25, 2013, 8 pgs.

International Search Report and Written Opinion for International Application No. PCT/US13/59991, Search Completed Feb. 6, 2014, 8 pgs.

(56) **References Cited**

OTHER PUBLICATIONS

International Search Report and Written Opinion for International Application No. PCT/US2009/044687, date completed Jan. 5, 2010, 9 pgs.
International Search Report and Written Opinion for International Application No. PCT/US2013/024987, Search Completed Mar. 27, 2013, 14 pgs.
International Search Report and Written Opinion for International Application No. PCT/US2013/056502, Search Completed Feb. 18, 2014, 7 pgs.
International Search Report and Written Opinion for International Application No. PCT/US2013/069932, Search Completed Mar. 14, 2014, 12 pgs.
International Search Report and Written Opinion for International Application PCT/US13/62720, report completed Mar. 25, 2014, 9 Pgs.
International Search Report and Written Opinion for International Application PCT/US14/024903 report completed Jun. 12, 2014, 13 pgs.
International Search Report and Written Opinion for International Application PCT/US14/17766, report completed May 28, 2014, 9 Pgs.
International Search Report and Written Opinion for International Application PCT/US14/18084, report completed May 23, 2014, 12 Pgs.
International Search Report and Written Opinion for International Application PCT/US14/18116, report completed May 13, 2014, 6 Pgs.
International Search Report and Written Opinion for International Application PCT/US14/22774 report completed Jun. 9, 2014, 6 Pgs.
International Search Report and Written Opinion for International Application PCT/US14/24407, report completed Jun. 11, 2014, 9 Pgs.
International Search Report and Written Opinion for International Application PCT/US14/25100, report completed Jul. 7, 2014, 5 Pgs.
International Search Report and Written Opinion for International Application PCT/US14/25904 report completed Jun. 10, 2014, 6 Pgs.
International Search Report and Written Opinion for International Application PCT/US2014/022123, report completed Jun. 9, 2014, 5 pgs.
International Search Report and Written Opinion for International Application PCT/US2014/024947, Report Completed Jul. 8, 2014, 8 Pgs.
International Search Report and Written Opinion for International Application PCT/US2014/028447, report completed Jun. 30, 2014, 8 Pgs.
International Search Report and Written Opinion for International Application PCT/US2014/030692, report completed Jul. 28, 2014, 7 Pages.
International Search Report and Written Opinion for International Application PCT/US2014/23762, Report Completed May 30, 2014, 6 Pgs.
IPRP for International Application No. PCT/US2012/089813, International Filing Date Oct. 11, 2012, Search Completed Apr. 15, 2014, 7 pgs.
Search Report and Written Opinion for International Application PCT/US11/36349, mailed Aug. 22, 2011, 12 pgs.
International Preliminary Report on Patentability for International Application PCT/US2013/039155, report completed Jul. 1, 2013, 10 Pgs.
International Search Report and Written Opinion for International Application No. PCT/US2011/64921, Report Completed Feb. 25, 2011, 17 pgs.
International Search Report and Written Opinion for International Application No. PCT/US2013/027146, completed Apr. 2, 2013, 12 pgs.
International Search Report and Written Opinion for International Application PCT/US14/22118, completed Jun. 9, 2014, Mailed, Jun. 25, 2014, 5 pgs.
International Search Report and Written Opinion for International Application PCT/US2010/057661, completed Mar. 9, 2011, 14 pgs.
International Search Report and Written Opinion for International Application PCT/US2012/044014, completed Oct. 12, 2012, 15 pgs.
International Search Report and Written Opinion for International Application PCT/US2012/056151, completed Nov. 14, 2012, 10 pgs.
International Search Report and Written Opinion for International Application PCT/US2012/059813, completed Dec. 17, 2012, 8 pgs.
International Search Report and Written Opinion for International Application PCT/US2012/37670, Mailed Jul. 18, 2012, Completed Jul. 5, 2012, 9 pgs.
International Search Report and Written Opinion for International Application PCT/US2012/58093, completed Nov. 15, 2012, 12 pgs.
Office Action for U.S. Appl. No. 12/952,106, dated Aug. 16, 2012, 12 pgs.
Baker et al., "Limits on Super-Resolution and How to Break Them", IEEE Transactions on Pattern Analysis and Machine Intelligence, Sep. 2002, vol. 24, No. 9, pp. 1167-1183.
Bertero et al., "Super-resolution in computational imaging", Micron, 2003, vol. 34, Issues 6-7, 17 pgs.
Bishop et al., "Full-Resolution Depth Map Estimation from an Aliased Plenoptic Light Field", ACCV 2010, Part II, LNCS 6493, pp. 186-200.
Bishop et al., "Light Field Superresolution", Retrieved from 2009 http://home.eps.hw.ac.uk/~sz73/ICCP09/LightFieldSuperresolution.pdf, 9 pgs.
Bishop et al., "The Light Field Camera: Extended Depth of Field, Aliasing, and Superresolution", IEEE Transactions on Pattern Analysis and Machine Intelligence, May 2012, vol. 34, No. 5, pp. 972-986.
Borman, "Topics in Multiframe Superresolution Restoration", Thesis of Sean Borman, Apr. 2004, 282 pgs.
Borman et al, "Image Sequence Processing", Source unknown, Oct. 14, 2002, 81 pgs.
Borman et al., "Block-Matching Sub-Pixel Motion Estimation from Noisy, Under-Sampled Frames—An Empirical Performance Evaluation", Proc SPIE, Dec. 1998, 3653, 10 pgs.
Borman et al., "Image Resampling and Constraint Formulation for Multi-Frame Super-Resolution Restoration", Proc. SPIE, Jun. 2003, 5016, 12 pgs.
Borman et al., "Linear models for multi-frame super-resolution restoration under non-affine registration and spatially varying PSF", Proc. SPIE, May 2004, vol. 5299, 12 pgs.
Borman et al., "Nonlinear Prediction Methods for Estimation of Clique Weighting Parameters in NonGaussian Image Models", Proc. SPIE, 1998. 3459, 9 pgs.
Borman et al., "Simultaneous Multi-Frame MAP Super-Resolution Video Enhancement Using Spatio-Temporal Priors", Image Processing, 1999, ICIP 99 Proceedings, vol. 3, pp. 469-473.
Borman et al., "Super-Resolution from Image Sequences—A Review", Circuits & Systems, 1998, pp. 374-378.
Krishnamurthy et al., "Compression and Transmission of Depth Maps for Image-Based Rendering", Image Processing, 2001, pp. 828-831.
Kutulakos et al., "Occluding Contour Detection Using Affine Invariants and Purposive Viewpoint Control", Proc., CVPR 94, 8 pgs, 1994.
Lensvector, "How LensVector Autofocus Works", printed Nov. 2, 2012 from http://www.lensvector.com/overview.html, 1 pg, 2012.
Levoy, "Light Fields and Computational Imaging", IEEE Computer Society, Aug. 2006, pp. 46-55.
Levoy et al., "Light Field Rendering", Proc. ADM SIGGRAPH '96, pp. 1-12, 1996.
Li et al., "A Hybrid Camera for Motion Deblurring and Depth Map Super-Resolution," Jun. 23-28, 2008, IEEE Conference on Computer Vision and Pattern Recognition, 8 pgs. Retrieved from www.eecis.udel.edu/~jye/lab_research/08/deblur-feng.pdf on Feb. 5, 2014.
Liu et al., "Virtual View Reconstruction Using Temporal Information", 2012 IEEE International Conference on Multimedia and Expo, 2012, pp. 115-120.
Lo et al., "Stereoscopic 3D Copy & Paste", ACM Transactions on Graphics, vol. 29, No. 6, Article 147, Dec. 2010, pp. 147:1-147:10.
Muehlebach, "Camera Auto Exposure Control for VSLAM Applications", Studies on Mechatronics, Swiss Federal Institute of Technology Zurich, Autumn Term 2010 course, 67 pgs.

(56) **References Cited**

OTHER PUBLICATIONS

Nayar, "Computational Cameras: Redefining the Image", IEEE Computer Society, Aug. 2006, pp. 30-38.
Ng, "Digital Light Field Photography", Thesis, Jul. 2006, 203 pgs.
Ng et al., "Super-Resolution Image Restoration from Blurred Low-Resolution Images", Journal of Mathematical Imaging and Vision, 2005, vol. 23, pp. 367-378.
Nitta et al., "Image reconstruction for thin observation module by bound optics by using the iterative backprojection method", Applied Optics, May 1, 2006, vol. 45, No. 13, pp. 2893-2900.
Nomura et al., "Scene Collages and Flexible Camera Arrays", Proceedings of Eurographics Symposium on Rendering, 2007, 12 pgs.
Park et al., "Super-Resolution Image Reconstruction", IEEE Signal Processing Magazine, May 2003, pp. 21-36.
Pham et al., "Robust Super-Resolution without Regularization", Journal of Physics: Conference Series 124, 2008, pp. 1-19.
Polight, "Designing Imaging Products Using Reflowable Autofocus Lenses", http://www.polight.no/tunable-polymer-autofocus-lens-html--11.html, 2012.
Protter et al., "Generalizing the Nonlocal-Means to Super-Resolution Reconstruction", IEEE Transactions on Image Processing, Jan. 2009, vol. 18, No. 1, pp. 36-51.
Radtke et al., "Laser lithographic fabrication and characterization of a spherical artificial compound eye", Optics Express, Mar. 19, 2007, vol. 15, No. 6, pp. 3067-3077.
Rander et al., "Virtualized Reality: Constructing Time-Varying Virtual Worlds From Real World Events", Proc. of IEEE Visualization '97, Phoenix, Arizona, Oct. 19-24, 1997, pp. 277-283, 552.
Rhemann et al, "Fast Cost-Volume Filtering for Visual Correspondence and Beyond", IEEE Trans. Pattern Anal. Mach. Intell, 2013, vol. 35, No. 2, pp. 504-511.
Robertson et al., "Dynamic Range Improvement Through Multiple Exposures", In Proc. of the Int. Conf. on Image Processing, 1999, 5 pgs.
Robertson et al., "Estimation-theoretic approach to dynamic range enhancement using multiple exposures", Journal of Electronic Imaging, Apr. 2003, vol. 12, No. 2, pp. 219-228.
Roy et al., "Non-Uniform Hierarchical Pyramid Stereo for Large Images", Computer and Robot Vision, 2007, pp. 208-215.
Sauer et al., "Parallel Computation of Sequential Pixel Updates in Statistical Tomographic Reconstruction", ICIP 1995, pp. 93-96.
Seitz et al., "Plenoptic Image Editing", International Journal of Computer Vision 48, 2, pp. 115-129, 2002.
Shum et al., "Pop-Up Light Field: An Interactive Image-Based Modeling and Rendering System," Apr. 2004, ACM Transactions on Graphics, vol. 23, No. 2, pp. 143-162. Retrieved from http://131.107.65.14/en-us/um/people/jiansun/papers/PopupLightField_TOG.pdf on Feb. 5.
Stollberg et al., "The Gabor superlens as an alternative wafer-level camera approach inspired by superposition compound eyes of nocturnal insects", Optics Express, Aug. 31, 2009, vol. 17, No. 18, pp. 15747-15759.
Sun et al., "Image Super-Resolution Using Gradient Profile Prior", Source and date unknown, 8 pgs, 2008.
Takeda et al., "Super-resolution Without Explicit Subpixel Motion Estimation", IEEE Transaction on Image Processing, Sep. 2009, vol. 18, No. 9, pp. 1958-1975.
Tanida et al., "Color imaging with an integrated compound imaging system", Optics Express, Sep. 8, 2003, vol. 11, No. 18, pp. 2109-2117.
Tanida et al., "Thin observation module by bound optics (TOMBO): concept and experimental verification", Applied Optics, Apr. 10, 2001, vol. 40, No. 11, pp. 1806-1813.
Taylor, "Virtual camera movement: The way of the future?", American Cinematographer 77, 9 (Sep.), 93-100, 1996.
Vaish et al., "Reconstructing Occluded Surfaces Using Synthetic Apertures: Stereo, Focus and Robust Measures", Proceeding, CVPR '06 Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition—vol. 2, pp. 2331-2338.
Vaish et al., "Synthetic Aperture Focusing Using a Shear-Warp Factorization of the Viewing Transform", IEEE Workshop on A3DISS, CVPR, 2005, 8 pgs.
Vaish et al., "Using Plane + Parallax for Calibrating Dense Camera Arrays", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004, 8 pgs.
Veilleux, "CCD Gain Lab: The Theory", University of Maryland, College Park-Observational Astronomy (ASTR 310), Oct. 19, 2006, pp. 1-5 (online), [retrieved on May 13, 2014]. Retrieved from the Internet <URL: http://www.astro.umd.edu/~veilleux/ASTR310/fall06/ccd_theory.pdf, 5 pgs.
Vuong et al., "A New Auto Exposure and Auto White-Balance Algorithm to Detect High Dynamic Range Conditions Using CMOS Technology", Proceedings of the World Congress on Engineering and Computer Science 2008, WCECS 2008, Oct. 22-24, 2008.
Wang, "Calculation of Image Position, Size and Orientation Using First Order Properties", 10 pgs, 2010.
Wetzstein et al., "Computational Plenoptic Imaging", Computer Graphics Forum, 2011, vol. 30, No. 8, pp. 2397-2426.
Wheeler et al., "Super-Resolution Image Synthesis Using Projections Onto Convex Sets in the Frequency Domain", Proc. SPIE, 2005, 5674, 12 pgs.
Wikipedia, "Polarizing Filter (Photography)", http://en.wikipedia.org/wiki/Polarizing_filter_(photography), 1 pg, 2012.
Wilburn, "High Performance Imaging Using Arrays of Inexpensive Cameras", Thesis of Bennett Wilburn, Dec. 2004, 128 pgs.
Wilburn et al., "High Performance Imaging Using Large Camera Arrays", ACM Transactions on Graphics, Jul. 2005, vol. 24, No. 3, pp. 765-776.
Wilburn et al., "High-Speed Videography Using a Dense Camera Array", Proceeding, CVPR'04 Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 294-301.
Wilburn et al., "The Light Field Video Camera", Proceedings of Media Processors 2002, SPIE Electronic Imaging, 2002, 8 pgs.
Wippermann et al., "Design and fabrication of a chirped array of refractive ellipsoidal micro-lenses for an apposition eye camera objective", Proceedings of SPIE, Optical Design and Engineering II, Oct. 15, 2005, 59622C-1-59622C-11.
Yang et al., "A Real-Time Distributed Light Field Camera", Eurographics Workshop on Rendering (2002), pp. 1-10.
Yang et al., "Superresolution Using Preconditioned Conjugate Gradient Method", Source and date unknown, 8 pgs. 2002.
Zhang et al., "A Self-Reconfigurable Camera Array", Eurographics Symposium on Rendering, 2004, 12 pgs.
Zomet et al., "Robust Super-Resolution", IEEE, 2001, pp. 1-6.

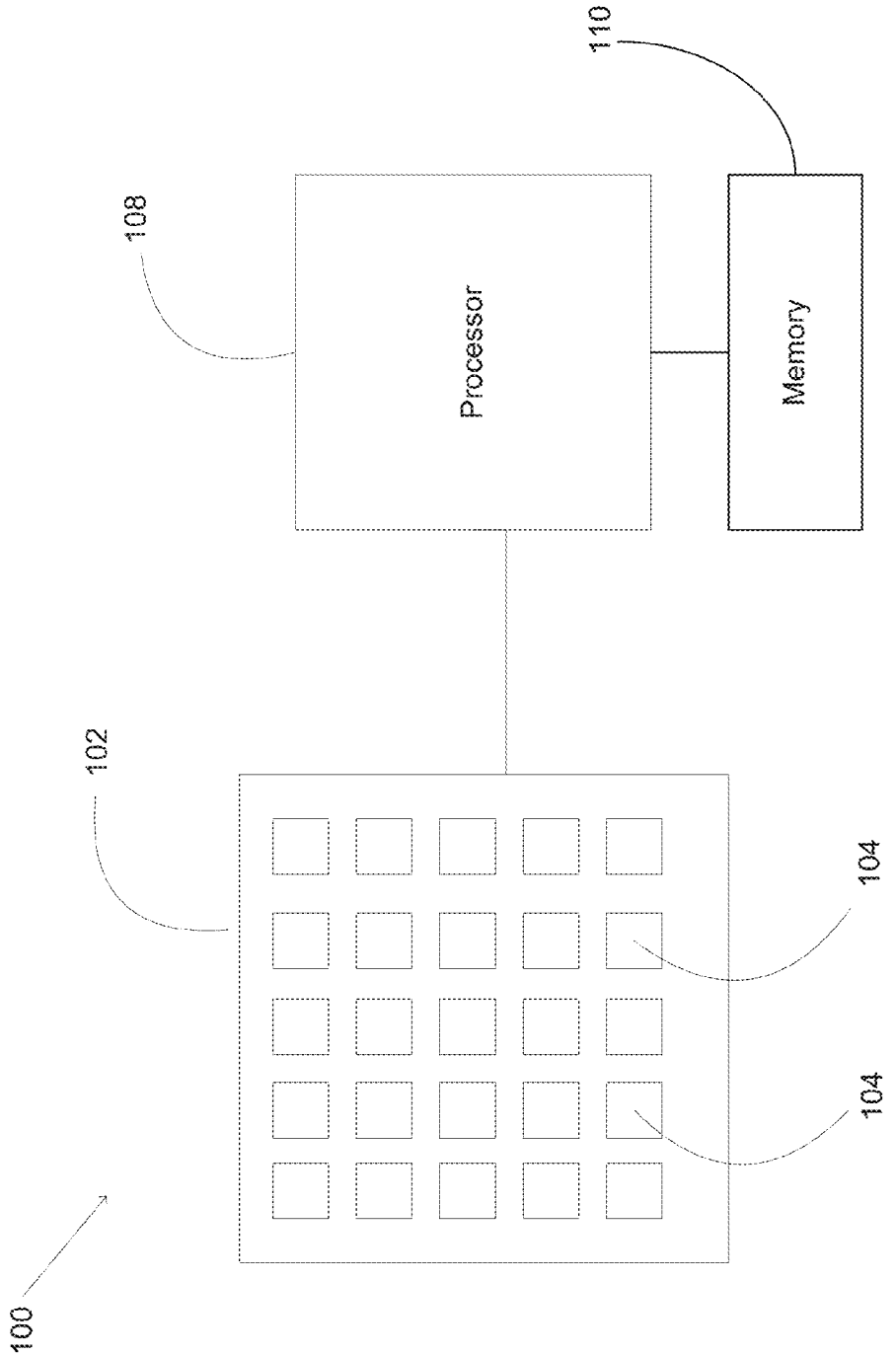


FIG. 1

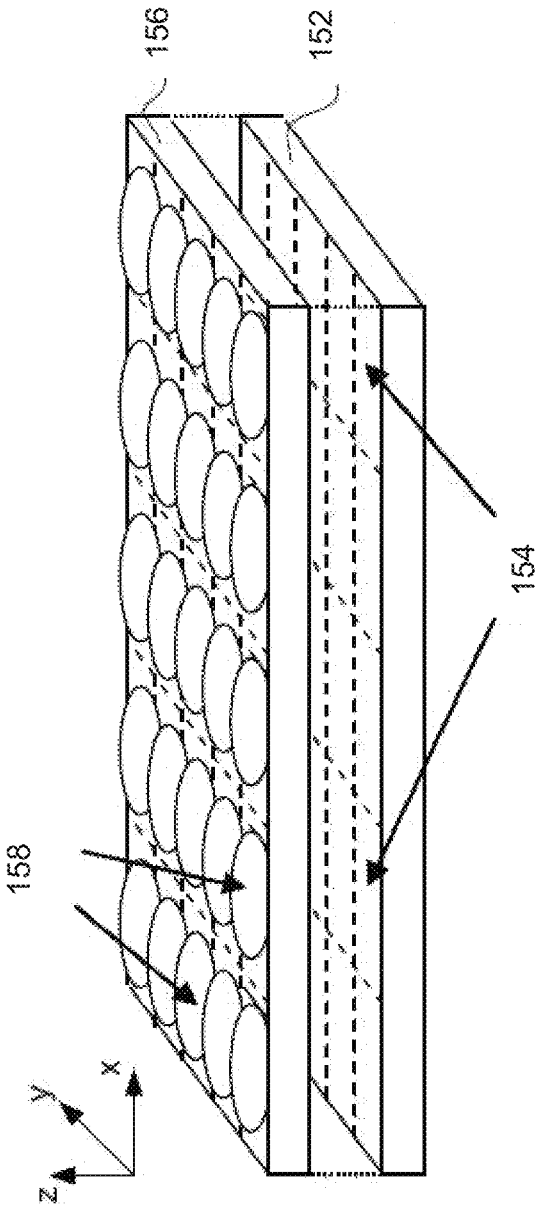


FIG. 1A

G	R	B	G
B	G	G	R
R	G	G	B
G	B	R	G

FIG. 1C

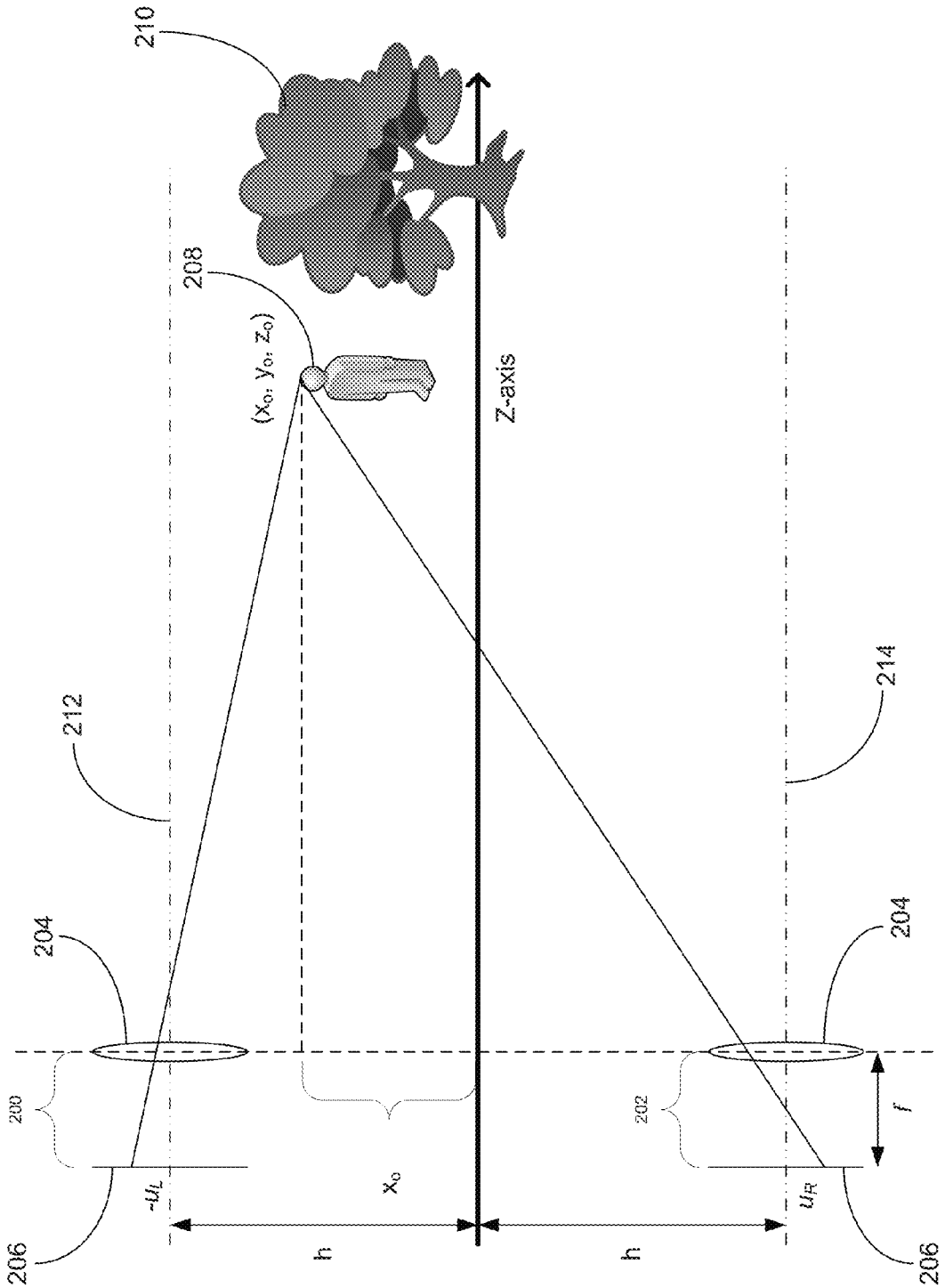

210
208
(x_o, y_o, z_o)
Z-axis
212
214
204
204
200
202
f
$-u_L$
x_o
u_R
206
206
h
h

FIG. 2

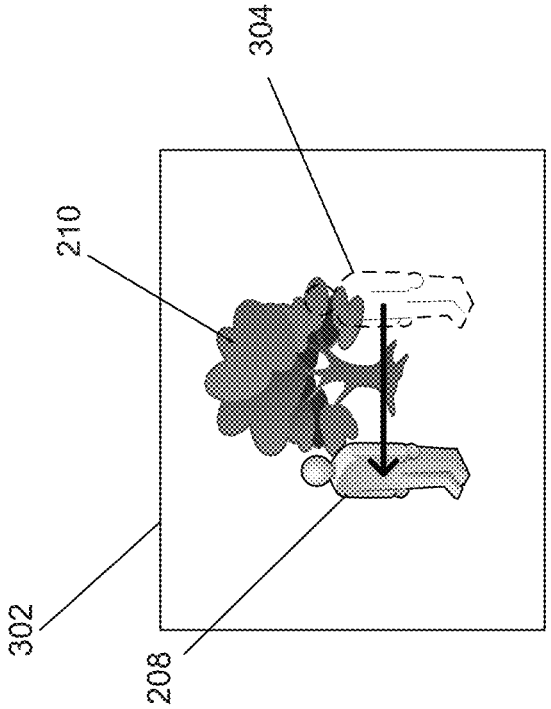


FIG. 3B

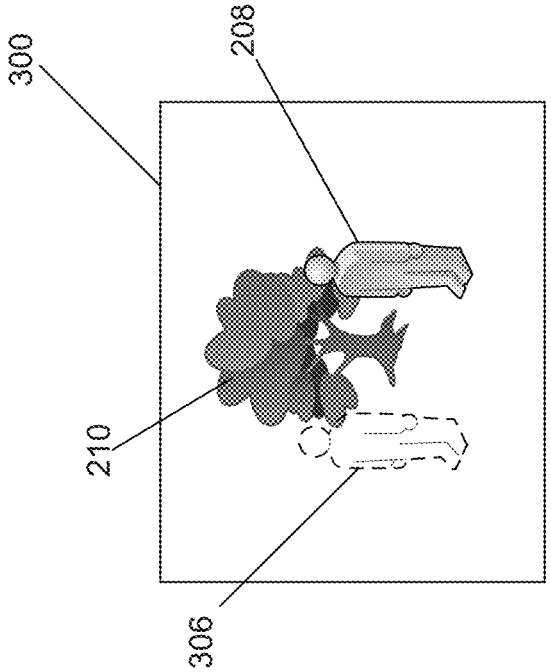


FIG. 3A

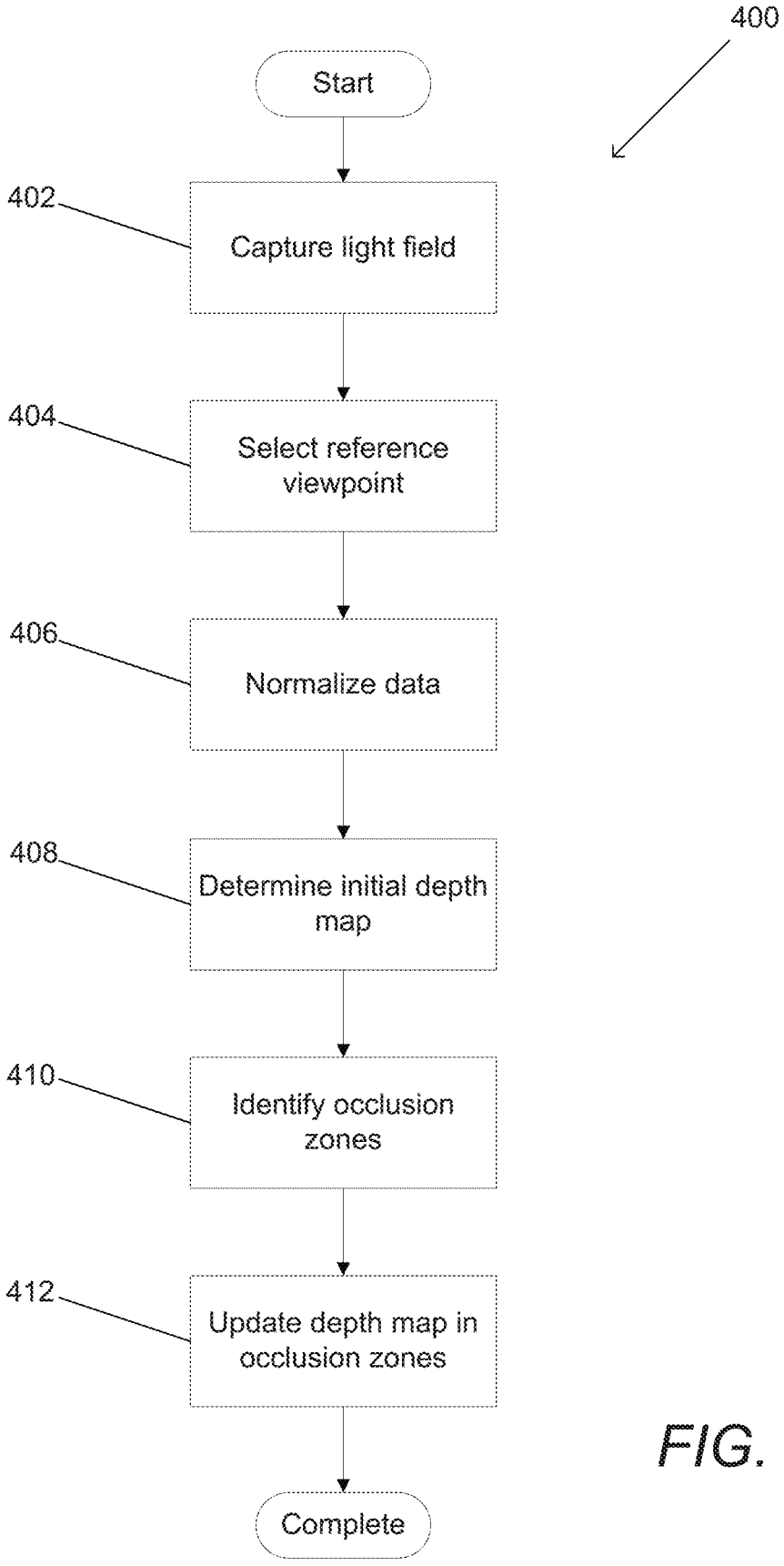

400
Start
402
Capture light field
404
Select reference viewpoint
406
Normalize data
408
Determine initial depth map
410
Identify occlusion zones
412
Update depth map in occlusion zones
Complete

FIG. 4

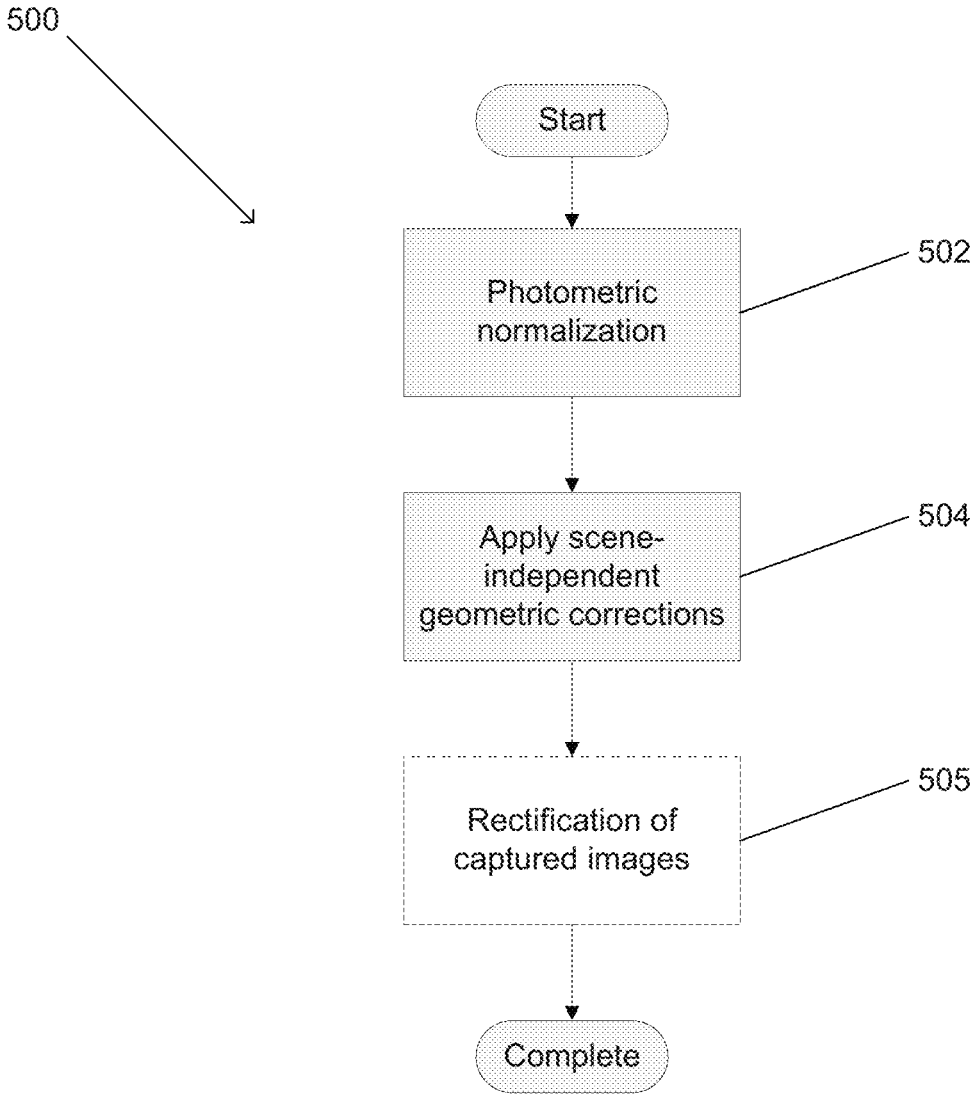

500
Start
Photometric normalization
502
Apply scene-independent geometric corrections
504
Rectification of captured images
505
Complete

FIG. 5

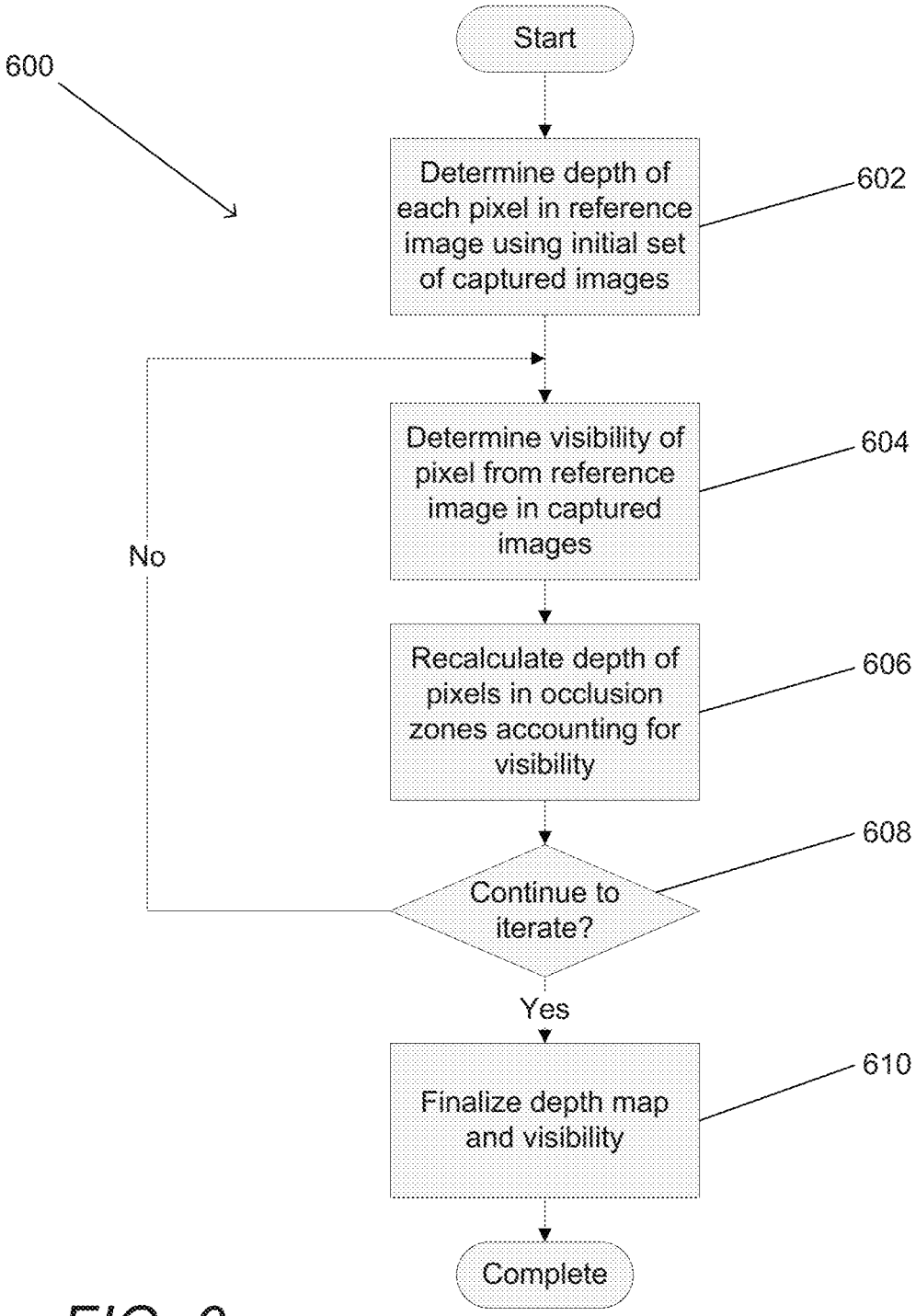

600
Start
Determine depth of each pixel in reference image using initial set of captured images
602
Determine visibility of pixel from reference image in captured images
604
No
Recalculate depth of pixels in occlusion zones accounting for visibility
606
Continue to iterate?
608
Yes
Finalize depth map and visibility
610
Complete

FIG. 6

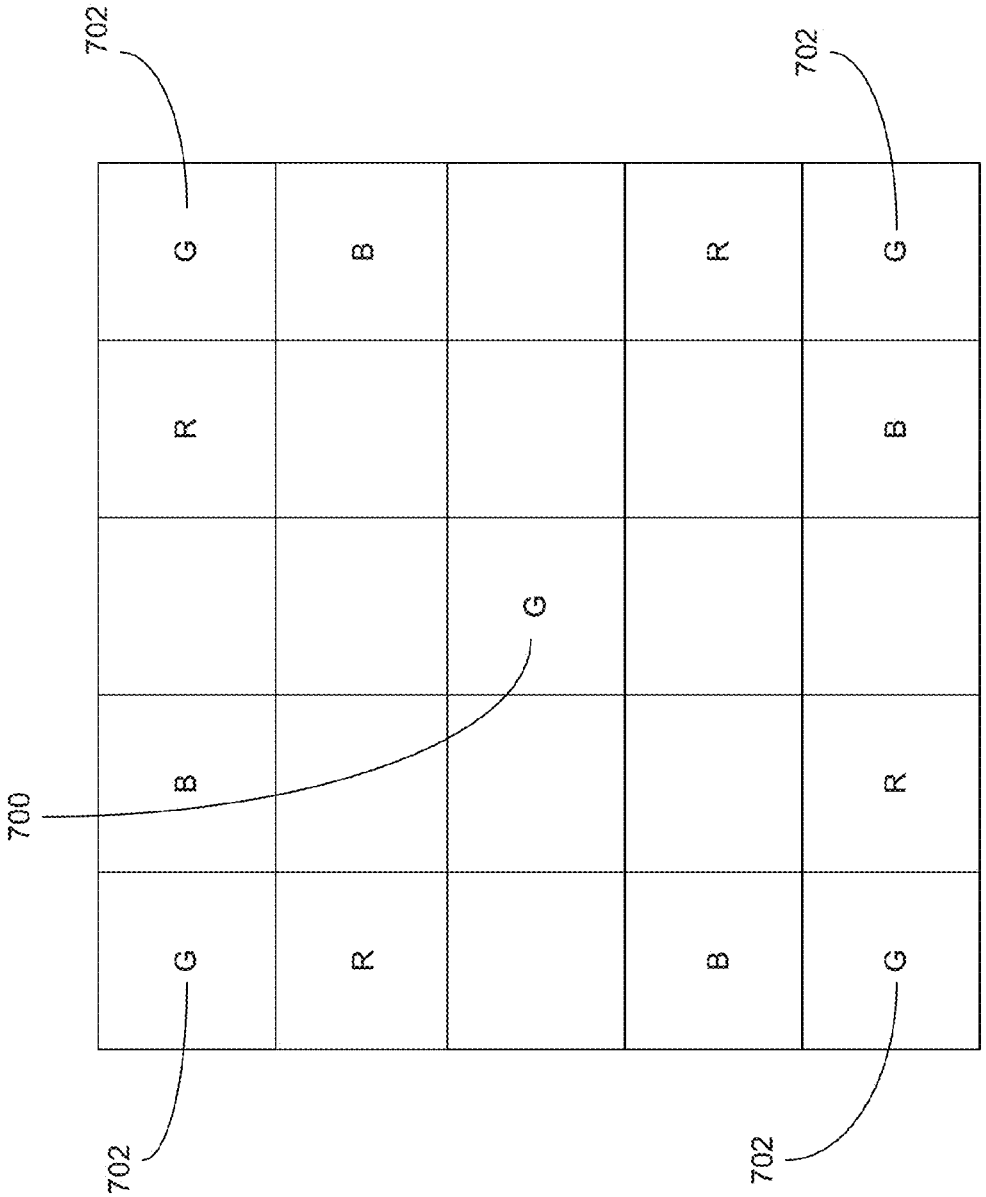


FIG. 7

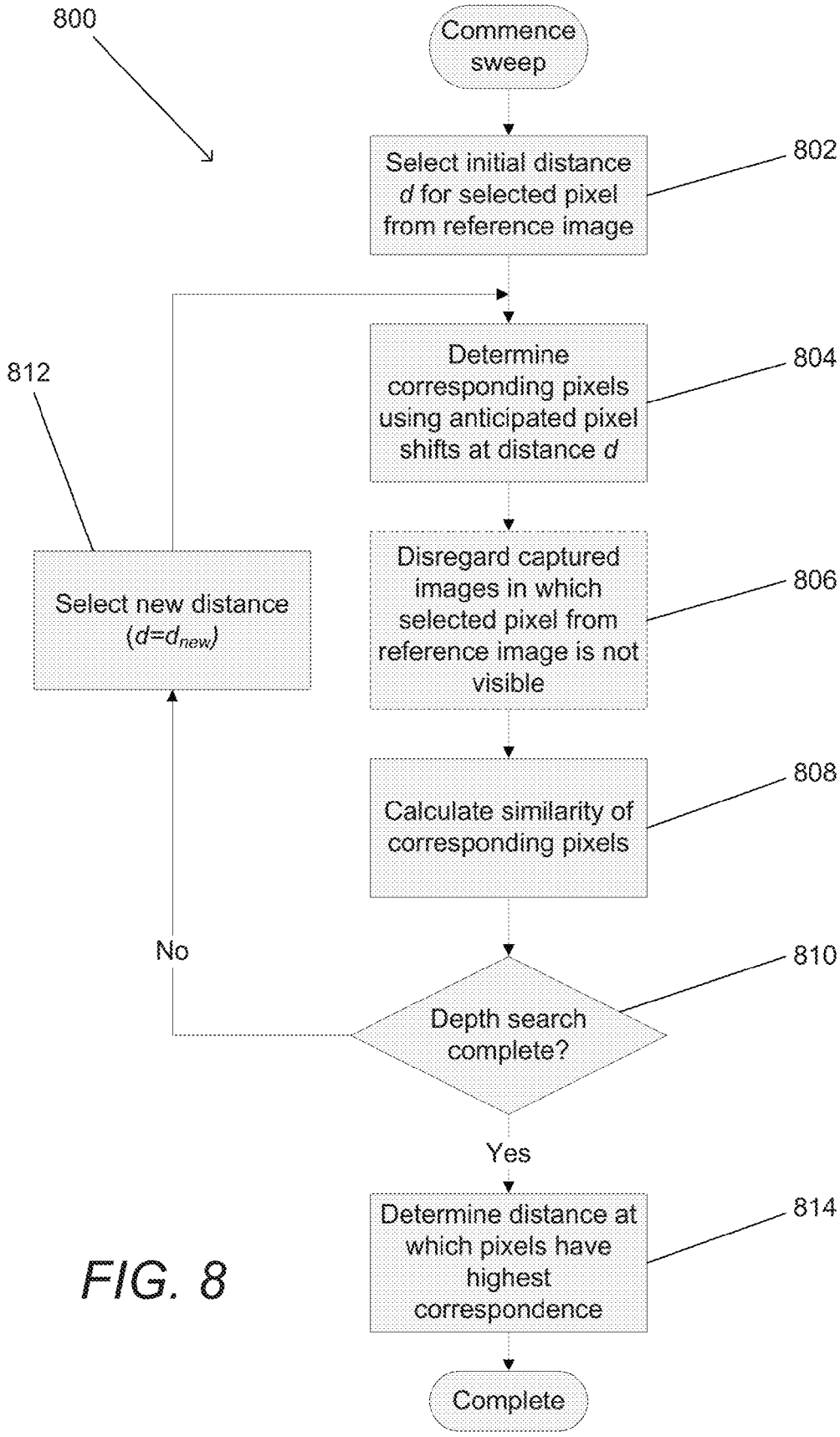

800
Commence sweep
Select initial distance d for selected pixel from reference image
802
Determine corresponding pixels using anticipated pixel shifts at distance d
804
Disregard captured images in which selected pixel from reference image is not visible
806
Calculate similarity of corresponding pixels
808
Depth search complete?
810
No
Yes
812
Select new distance (d=d_new)
Determine distance at which pixels have highest correspondence
814
Complete

FIG. 8

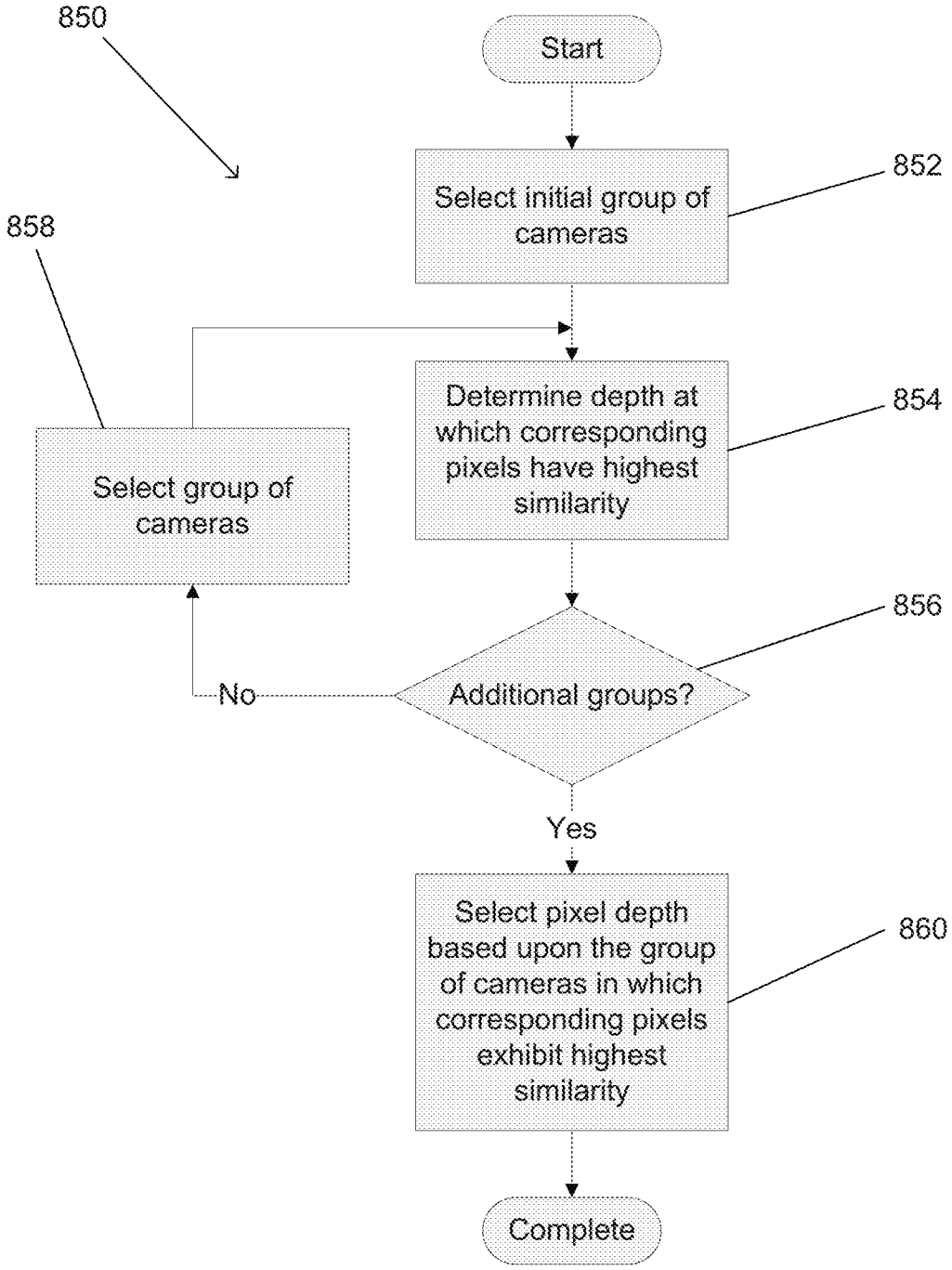

850
Start
Select initial group of cameras
852
Determine depth at which corresponding pixels have highest similarity
854
Additional groups?
856
No
Select group of cameras
858
Yes
Select pixel depth based upon the group of cameras in which corresponding pixels exhibit highest similarity
860
Complete

FIG. 8A

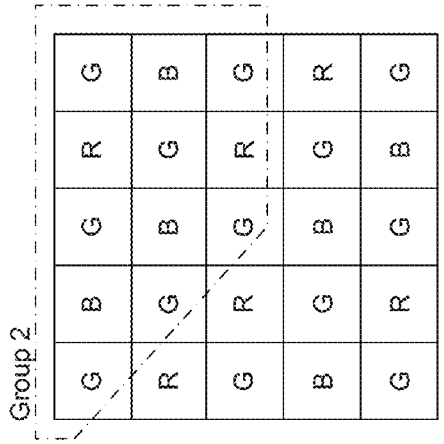


FIG. 8C

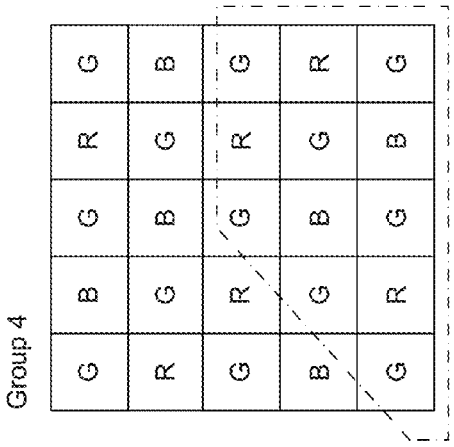


FIG. 8E

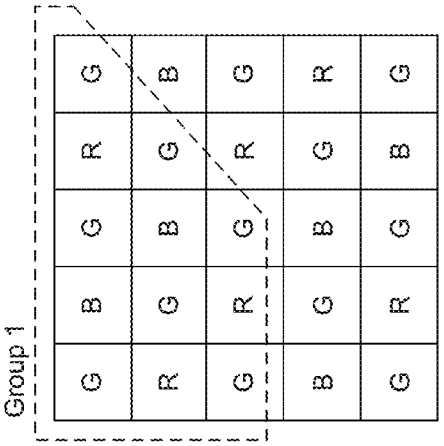


FIG. 8B

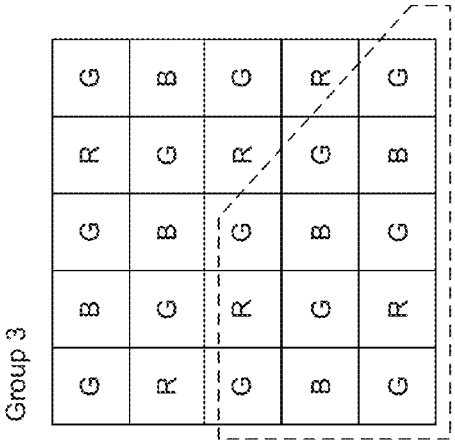


FIG. 8D

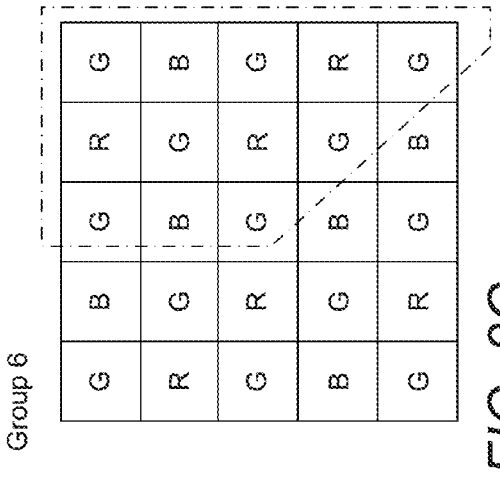


FIG. 8G

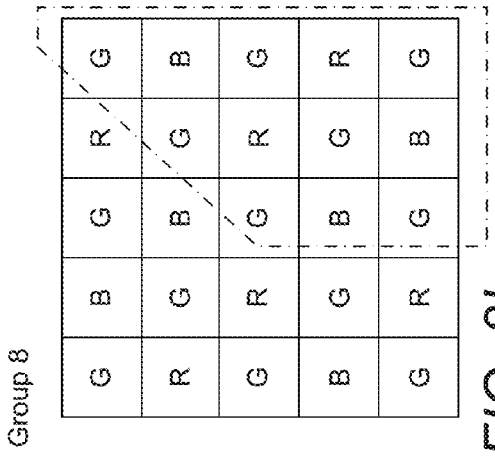


FIG. 8I

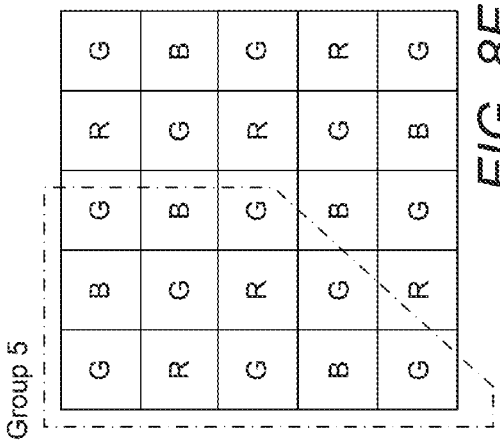


FIG. 8F

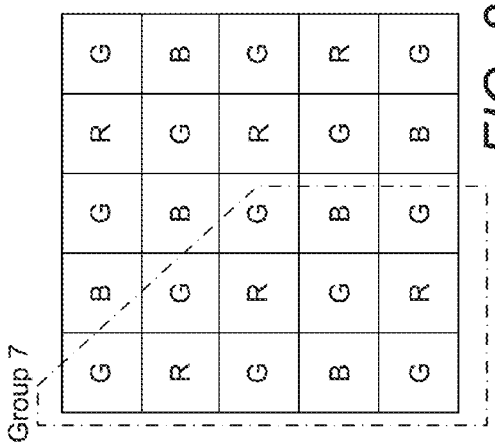


FIG. 8H

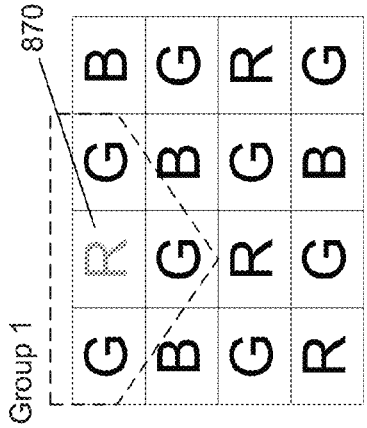


FIG. 8J

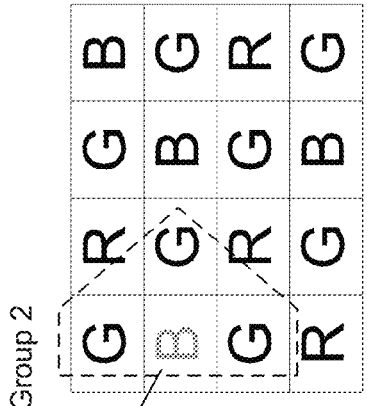


FIG. 8K

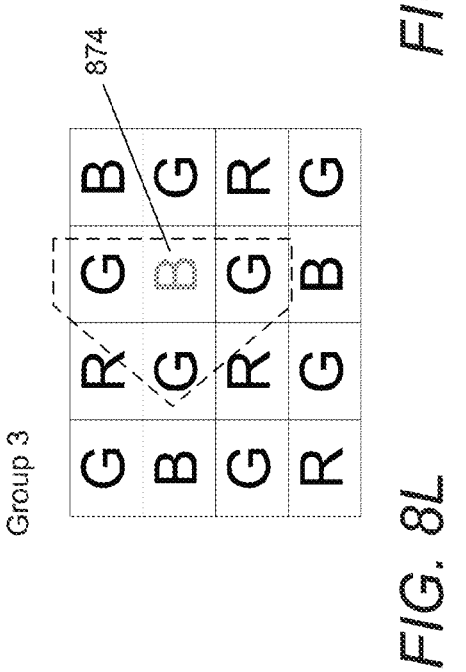


FIG. 8L

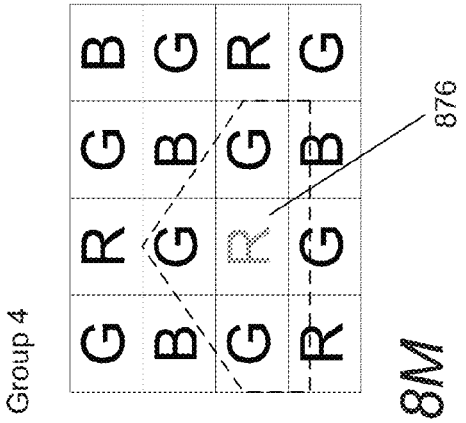


FIG. 8M

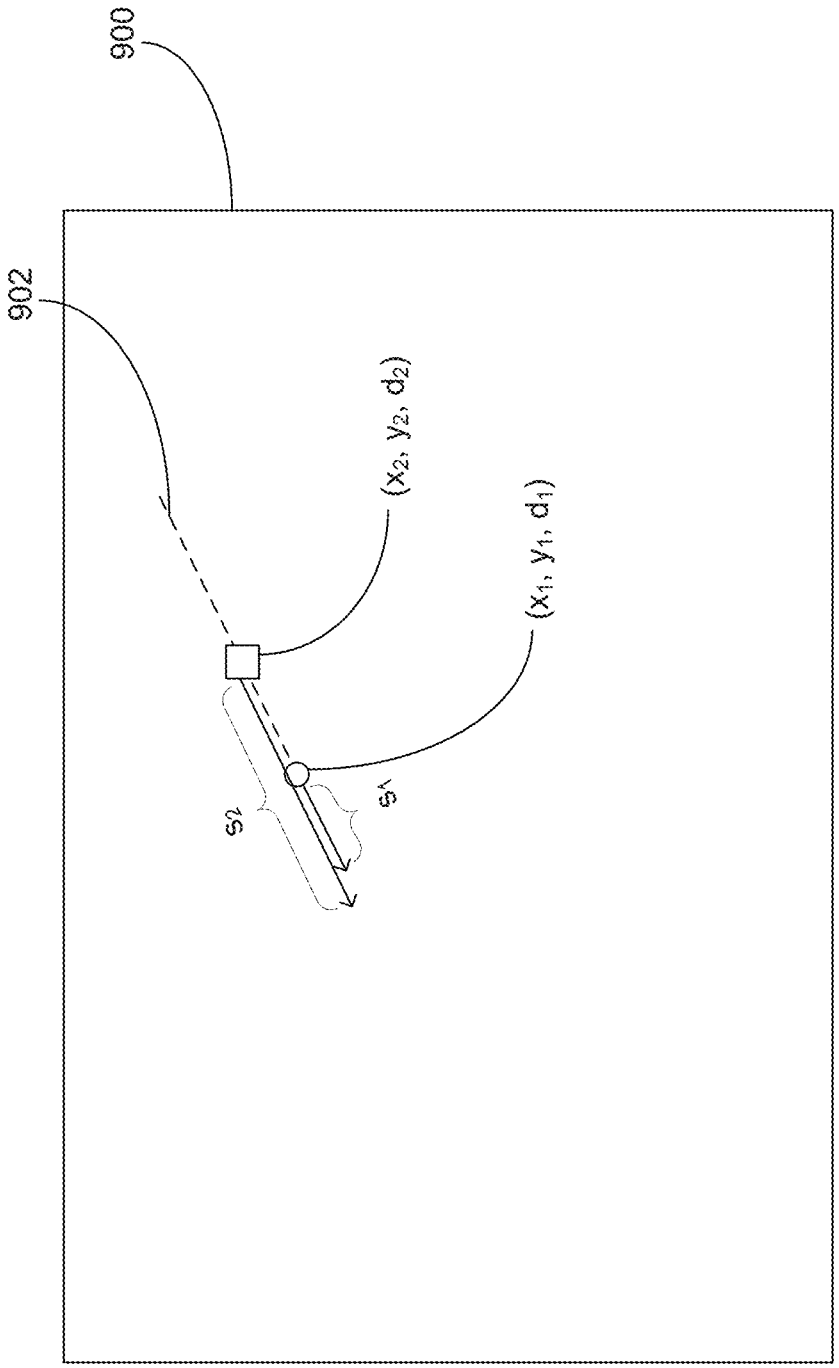


FIG. 9

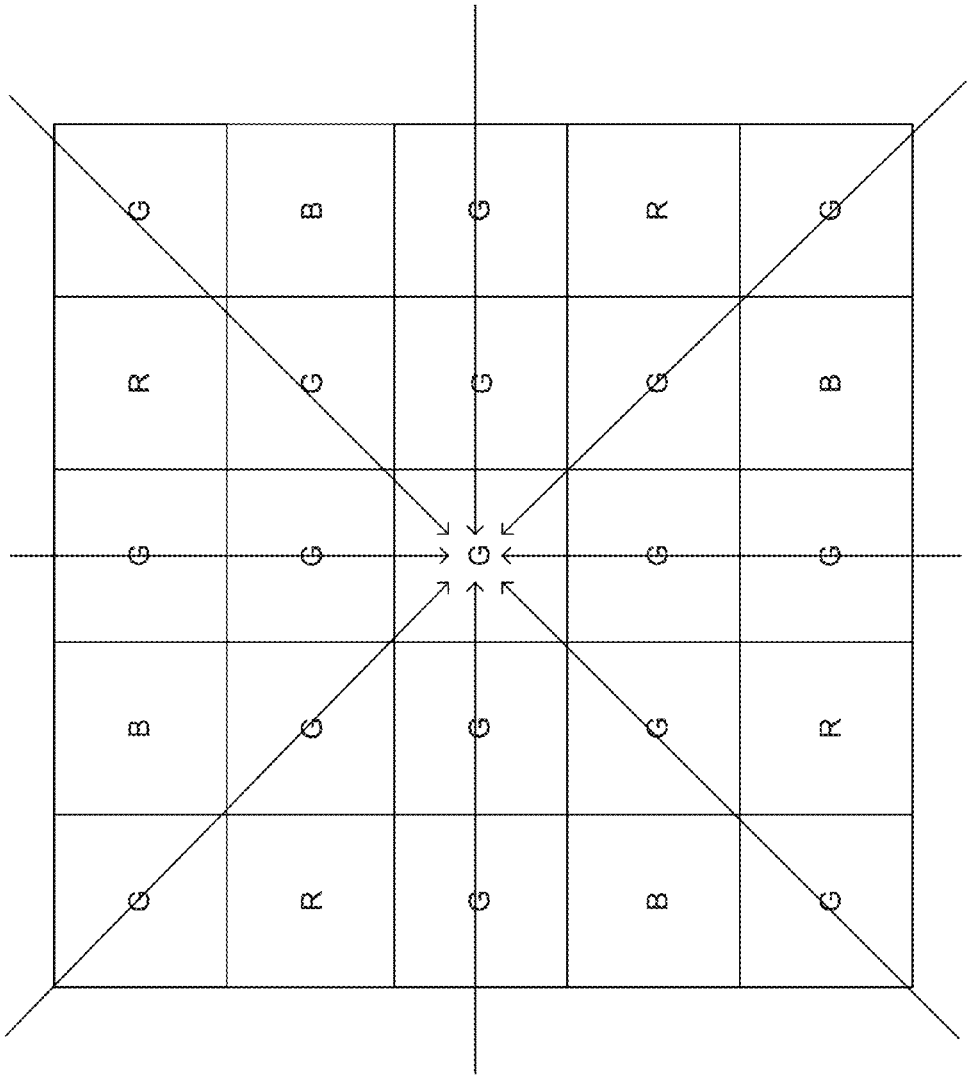


FIG. 10

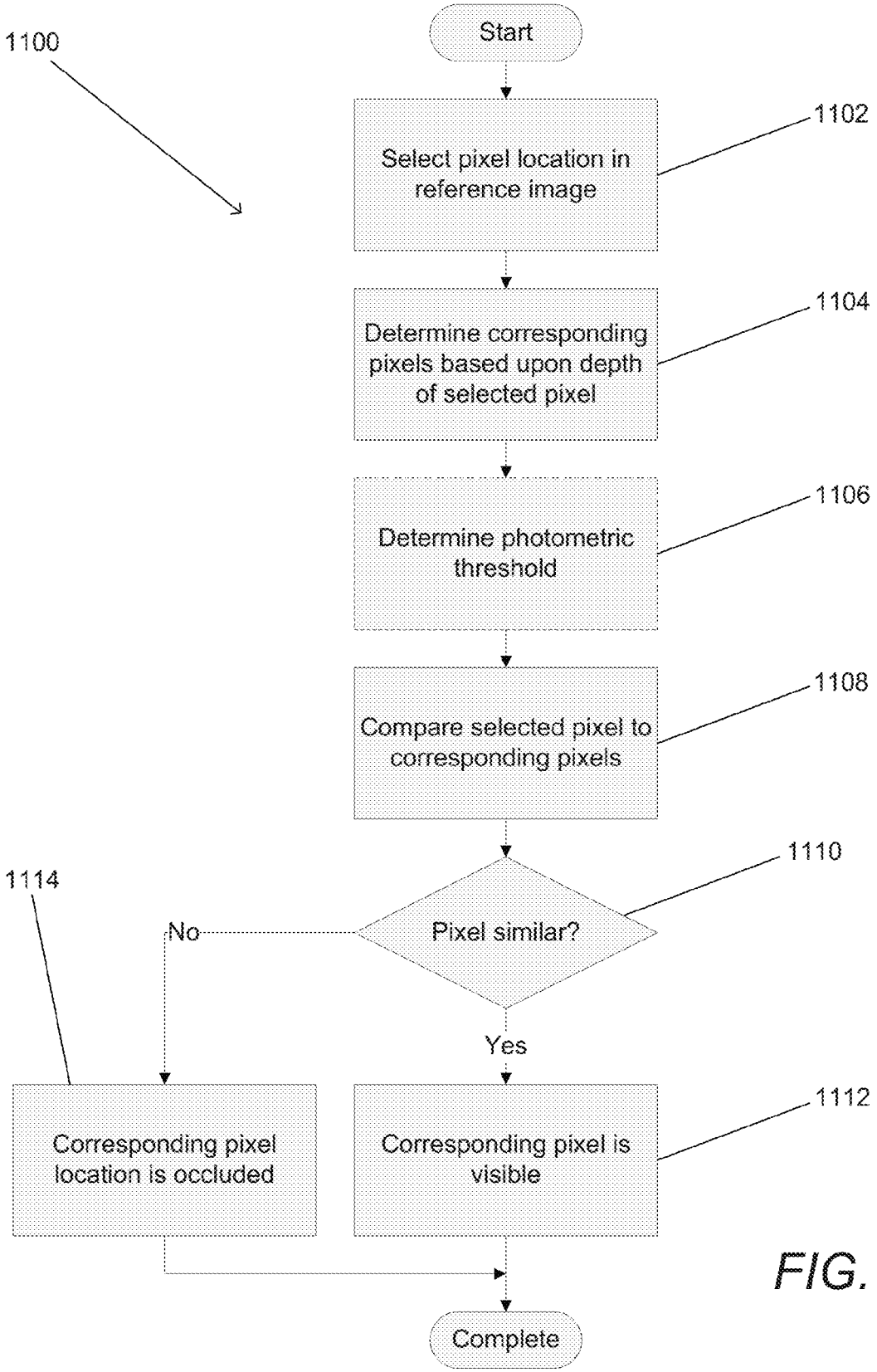

1100
Start
Select pixel location in reference image
1102
Determine corresponding pixels based upon depth of selected pixel
1104
Determine photometric threshold
1106
Compare selected pixel to corresponding pixels
1108
Pixel similar?
1110
No
Yes
1114
Corresponding pixel location is occluded
Corresponding pixel is visible
1112
Complete

FIG. 11

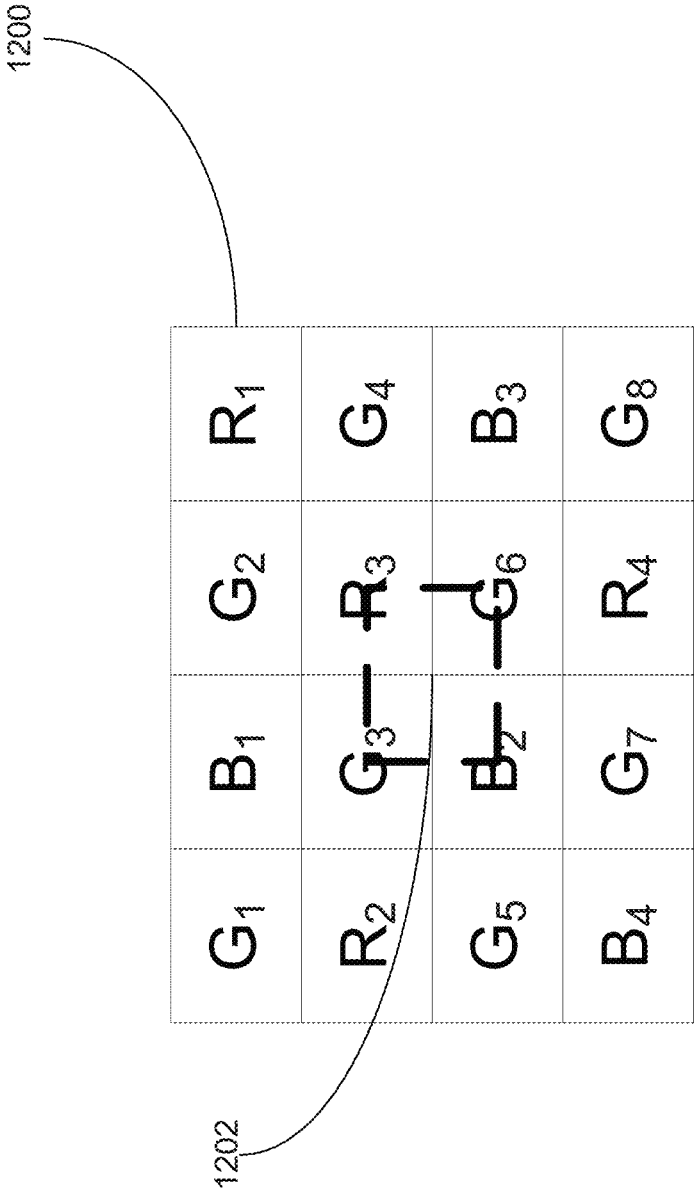


FIG. 12

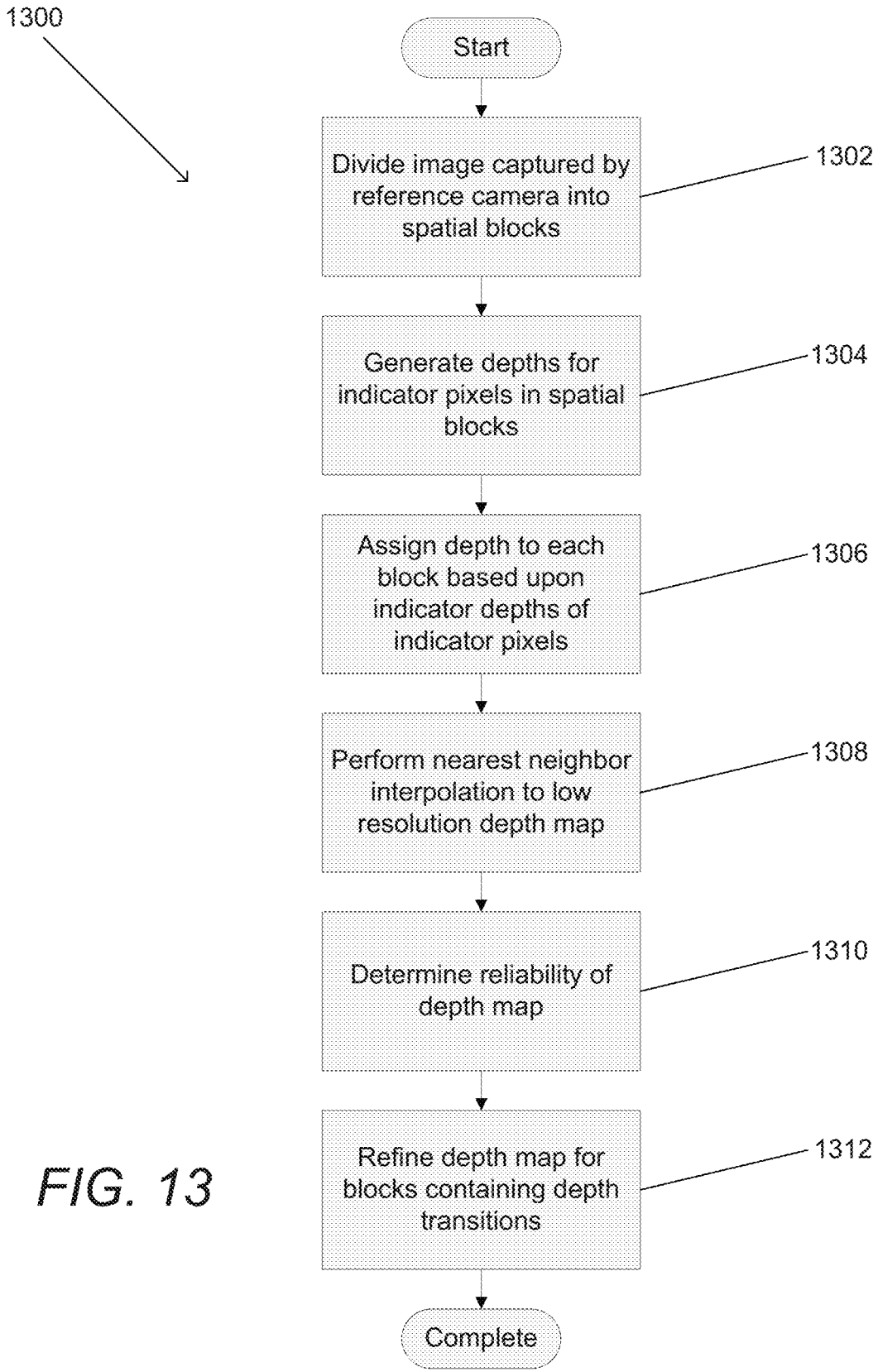

1300
Start
Divide image captured by reference camera into spatial blocks
1302
Generate depths for indicator pixels in spatial blocks
1304
Assign depth to each block based upon indicator depths of indicator pixels
1306
Perform nearest neighbor interpolation to low resolution depth map
1308
Determine reliability of depth map
1310
Refine depth map for blocks containing depth transitions
1312
Complete

FIG. 13

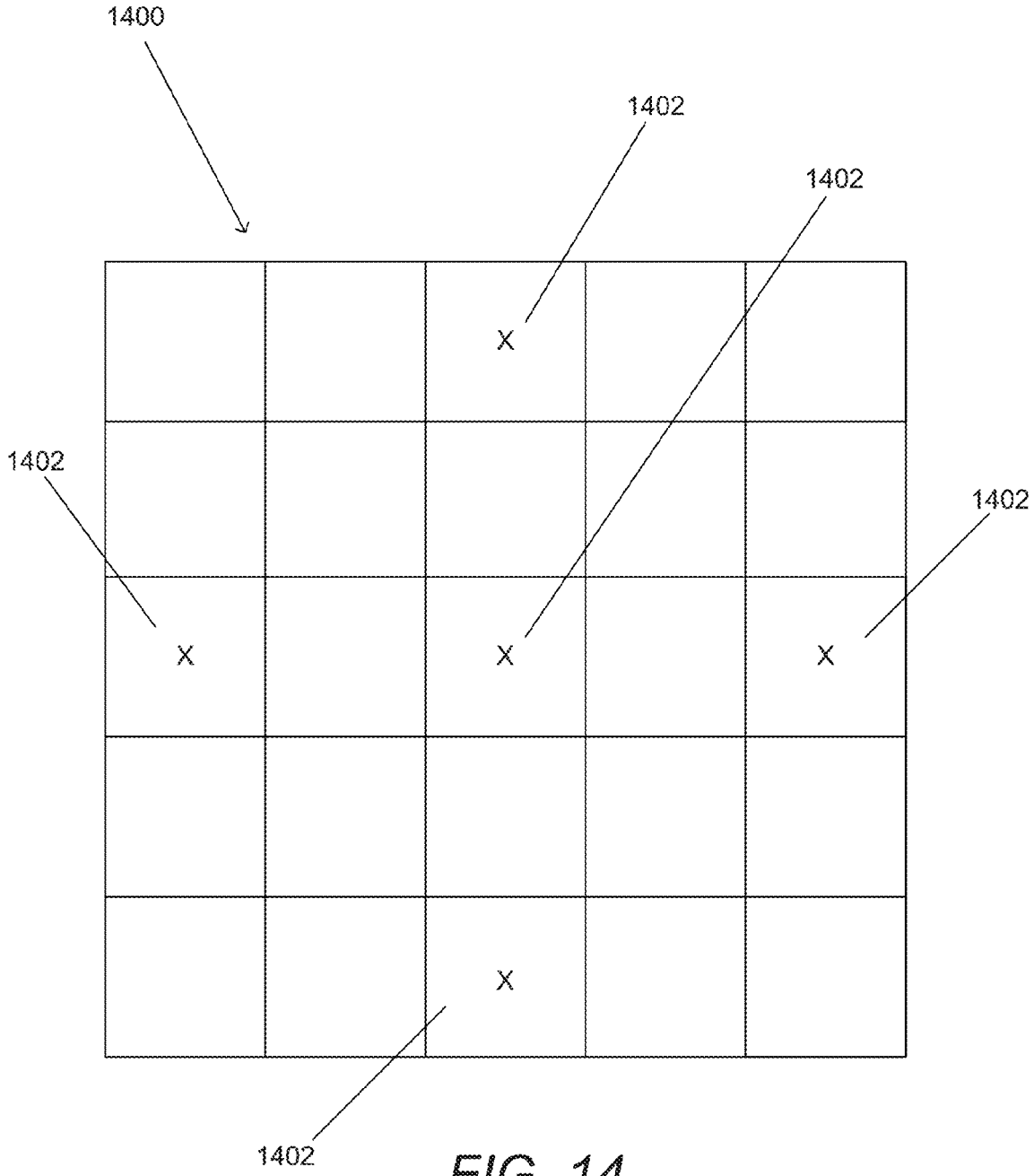


FIG. 14

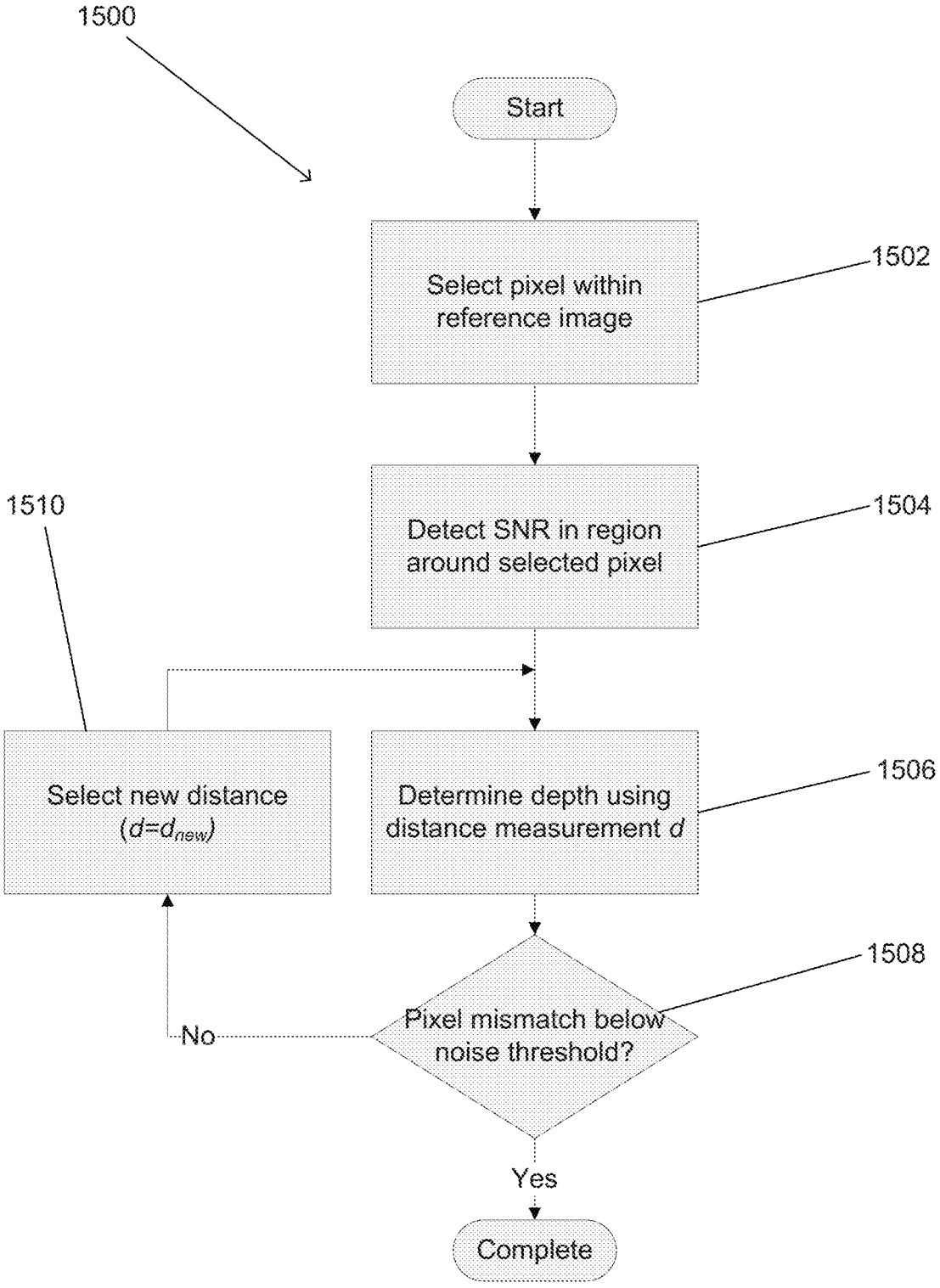

1500
Start
Select pixel within reference image
1502
Detect SNR in region around selected pixel
1504
1510
Select new distance ($d=d_{new}$)
Determine depth using distance measurement d
1506
Pixel mismatch below noise threshold?
1508
No
Yes
Complete

FIG. 15

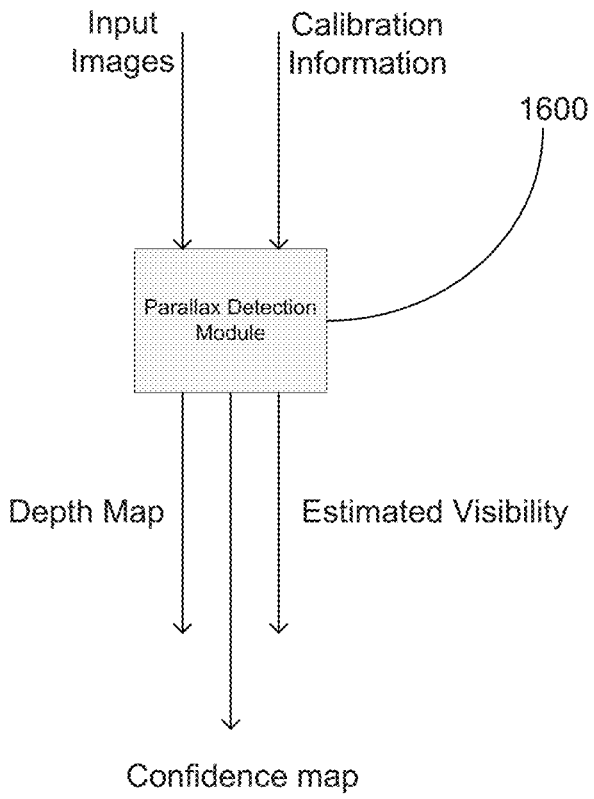

Input Images
Calibration Information
1600
Parallax Detection Module
Depth Map
Estimated Visibility
Confidence map

FIG. 16

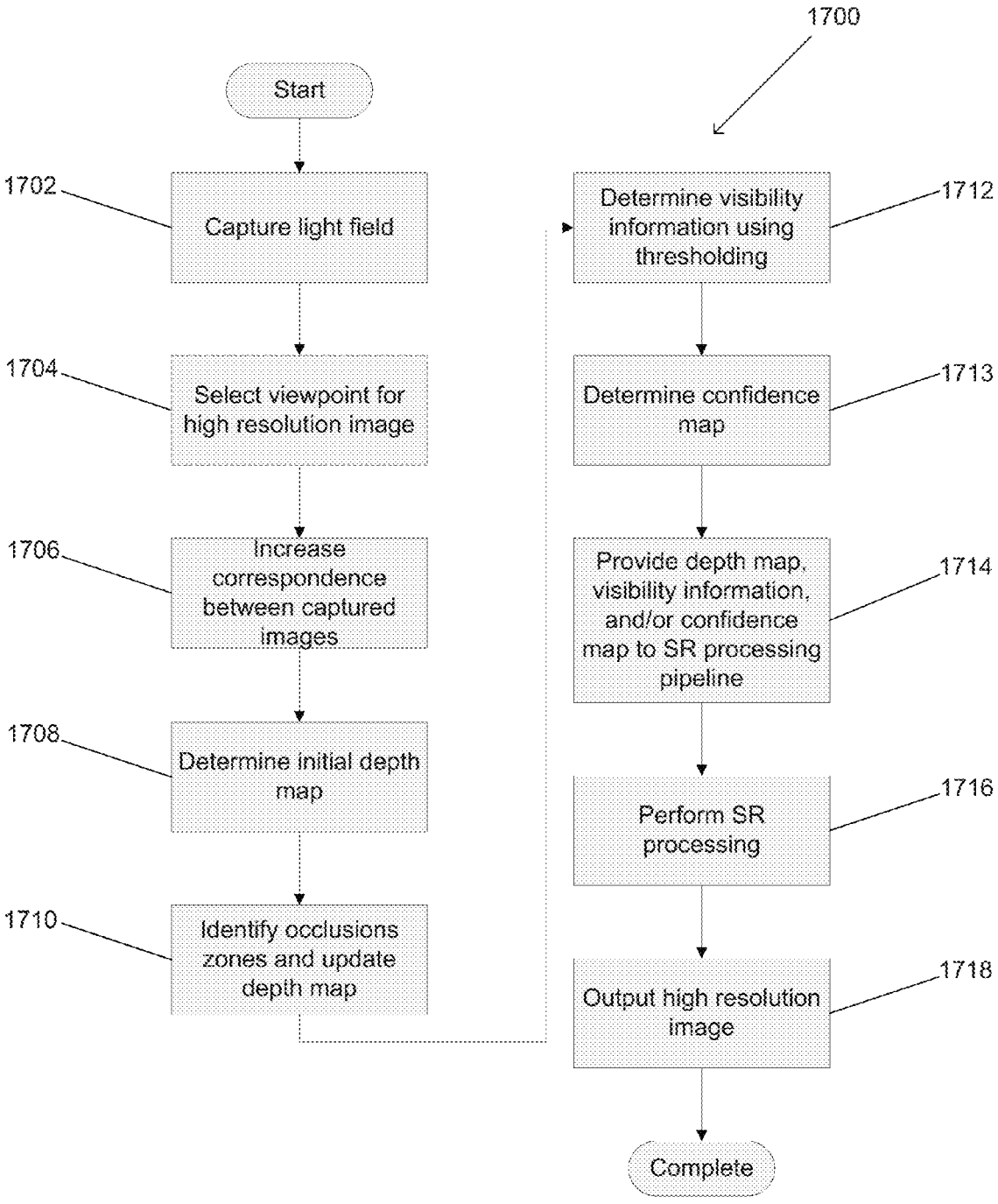

1700
Start
1702
Capture light field
1704
Select viewpoint for high resolution image
1706
Increase correspondence between captured images
1708
Determine initial depth map
1710
Identify occlusions zones and update depth map
Determine visibility information using thresholding
1712
Determine confidence map
1713
Provide depth map, visibility information, and/or confidence map to SR processing pipeline
1714
Perform SR processing
1716
Output high resolution image
1718
Complete

FIG. 17

FIG. 18A

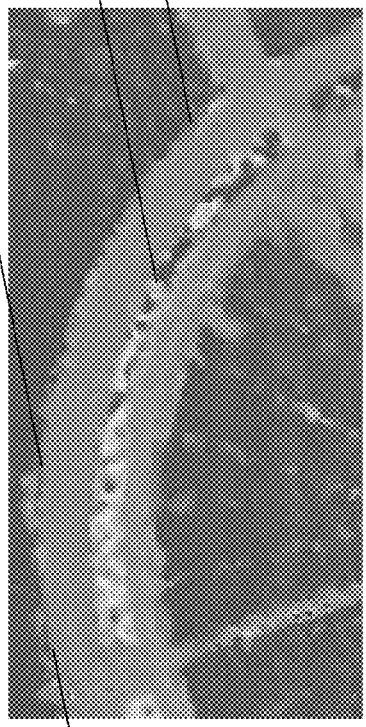


FIG. 18B

FIG. 18C

FIG. 18D

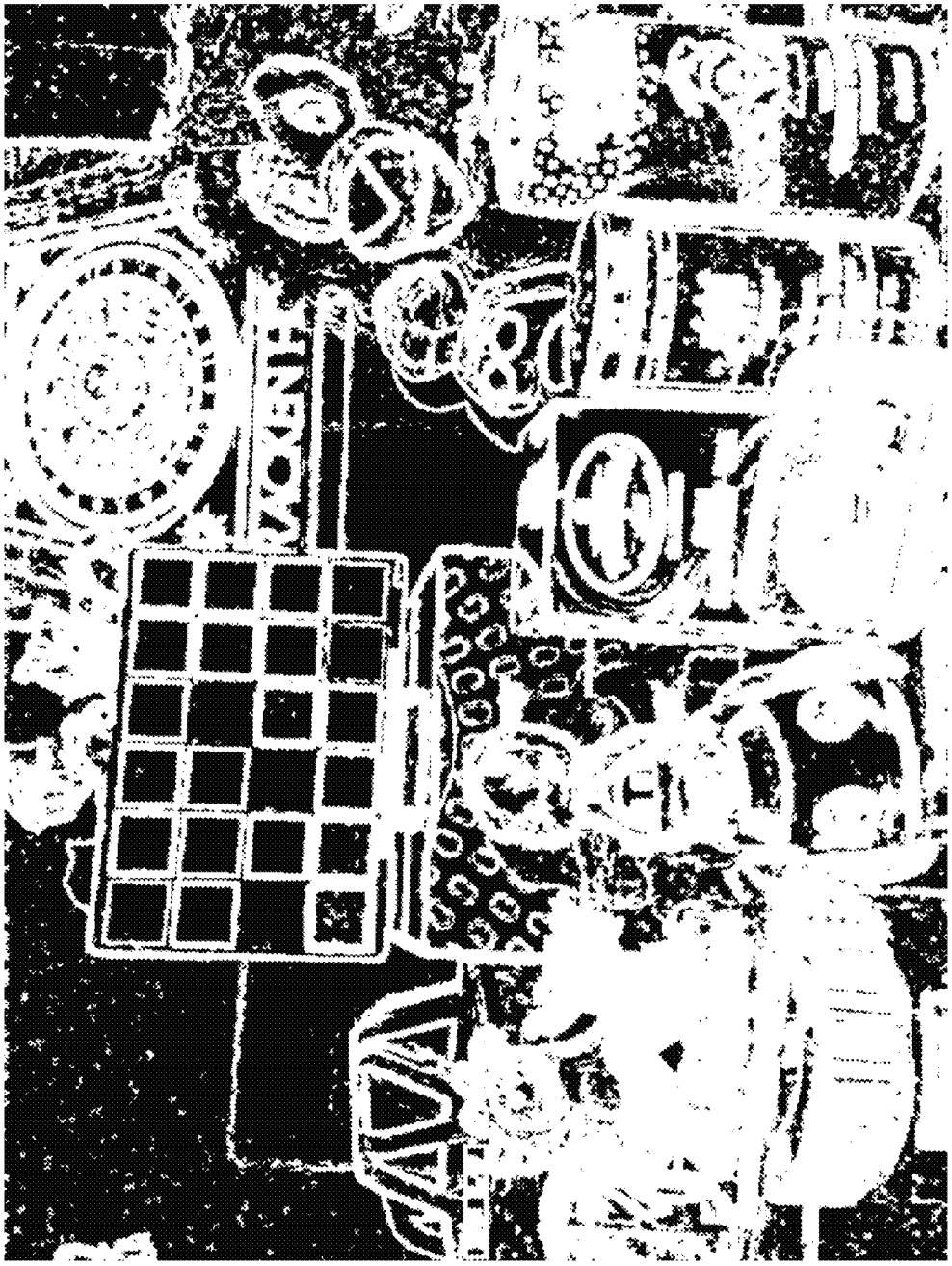


FIG. 18E

FIG. 18F

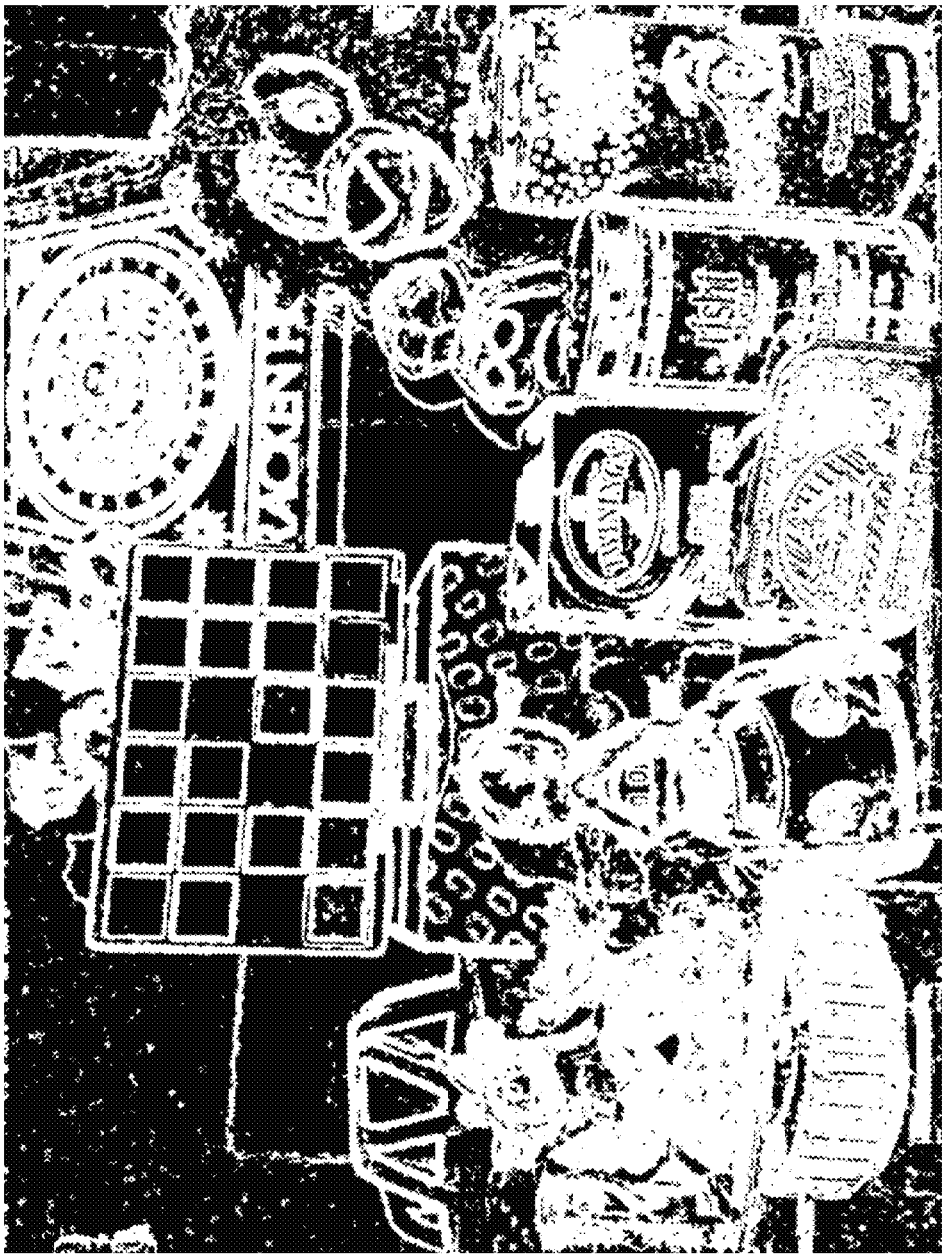


FIG. 18G

FIG. 18H

FIG. 18J

FIG. 18I

FIG. 18L

FIG. 18K

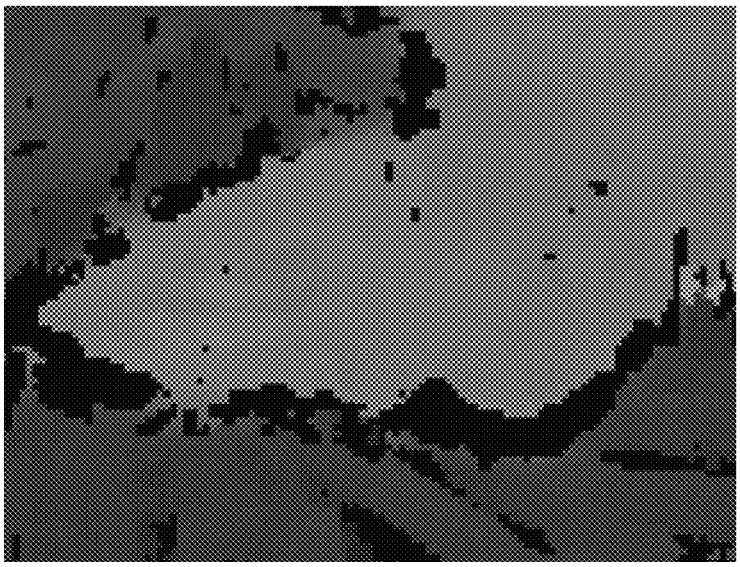


FIG. 18N

FIG. 18M

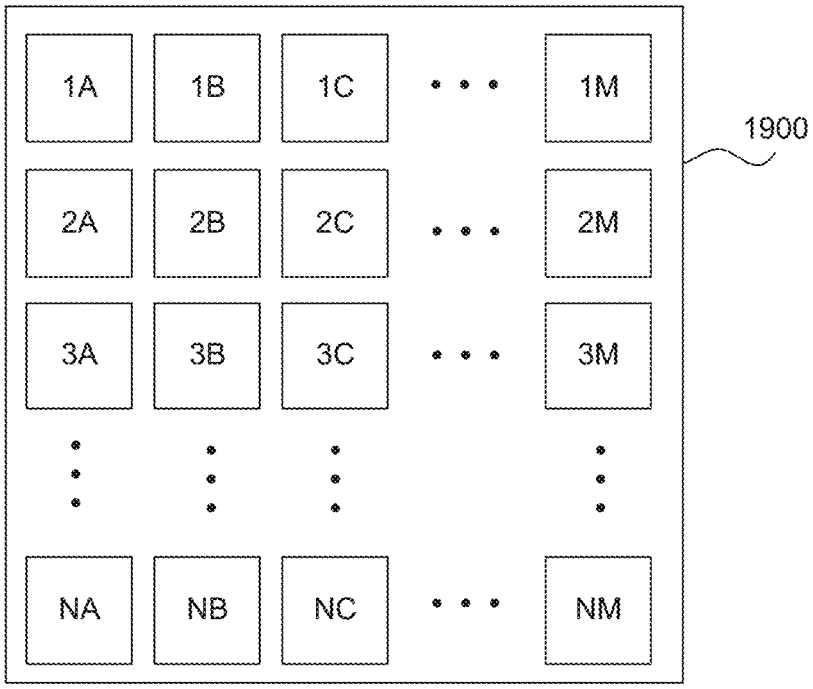


FIG. 19

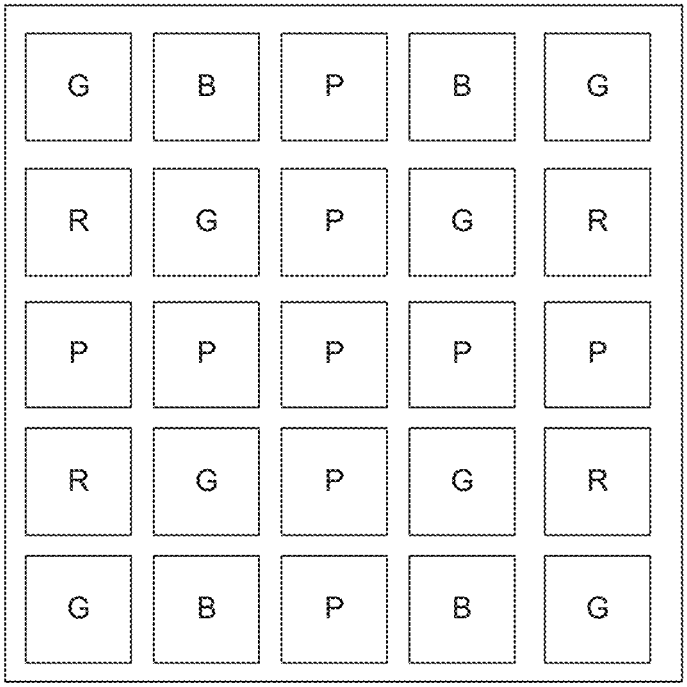


FIG. 20A

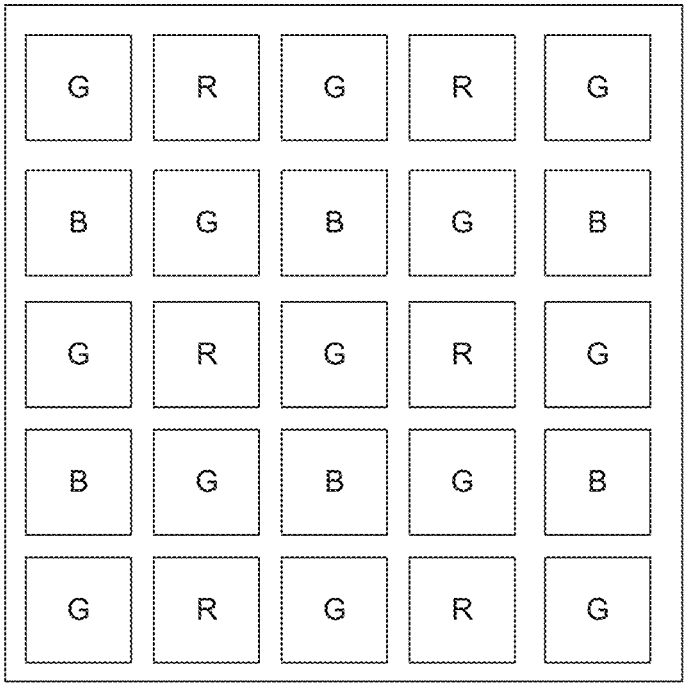


FIG. 20B

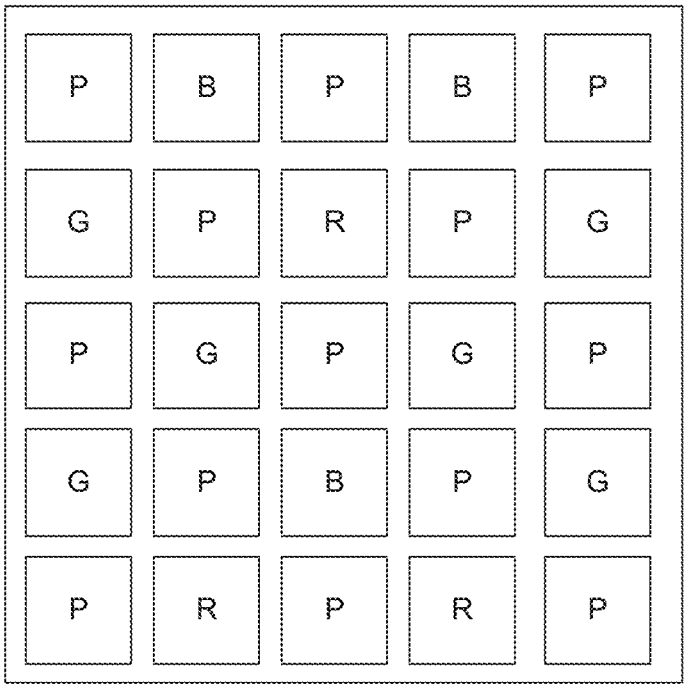


FIG. 20C

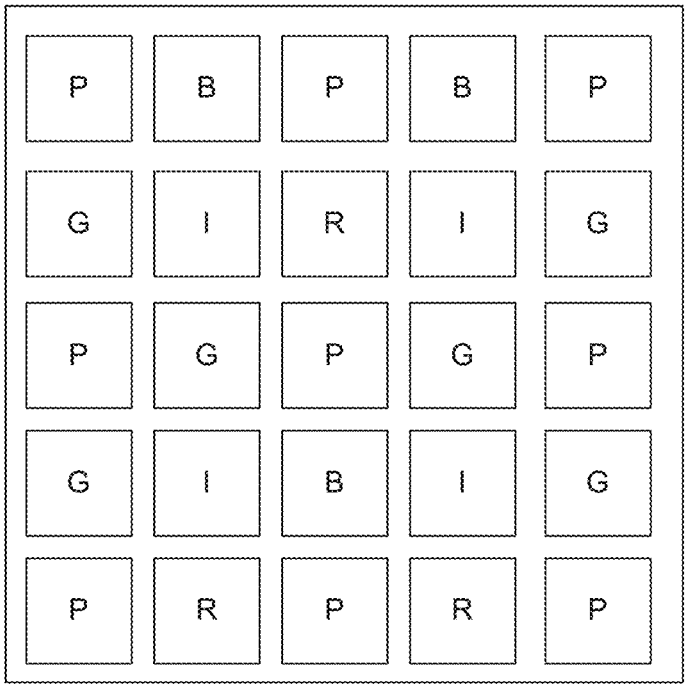


FIG. 20D

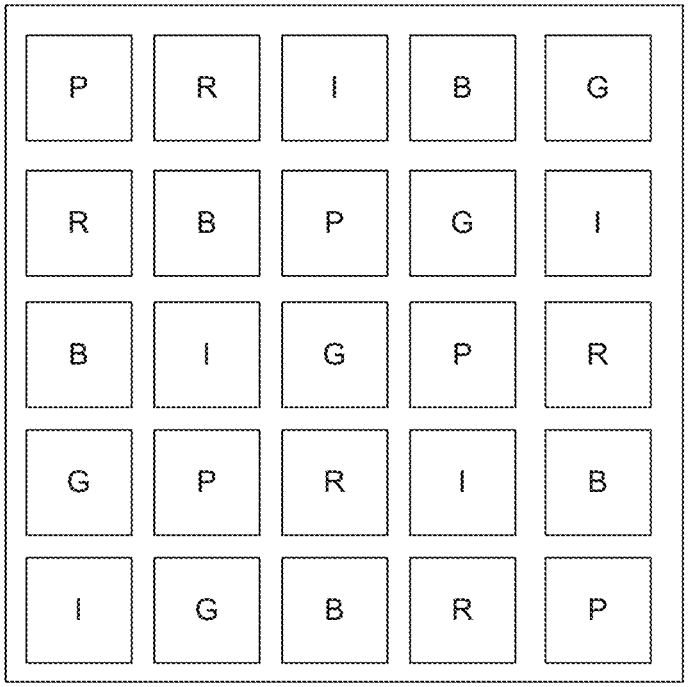


FIG. 20E

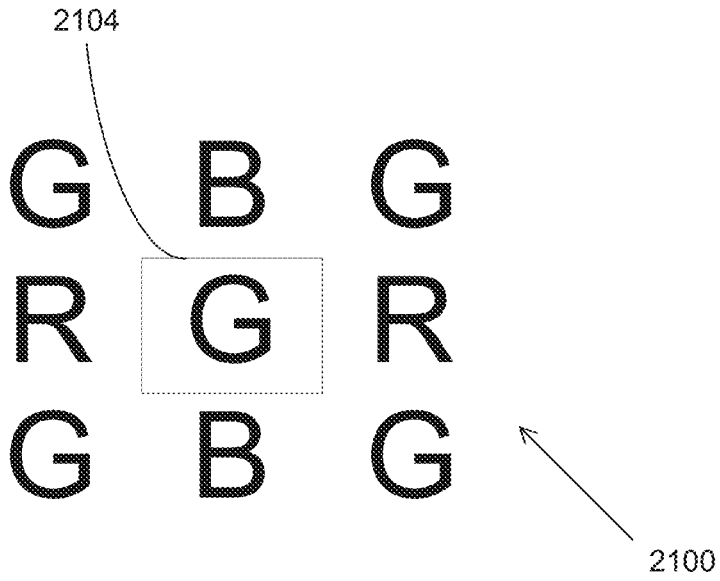


FIG. 21A

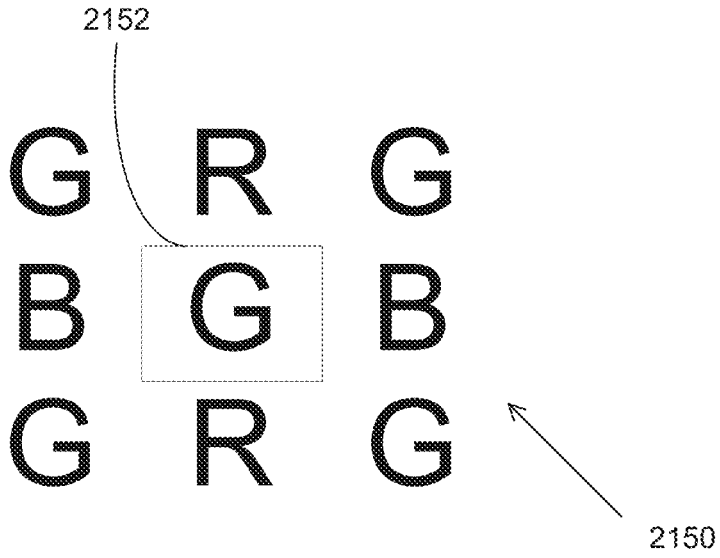


FIG. 21B

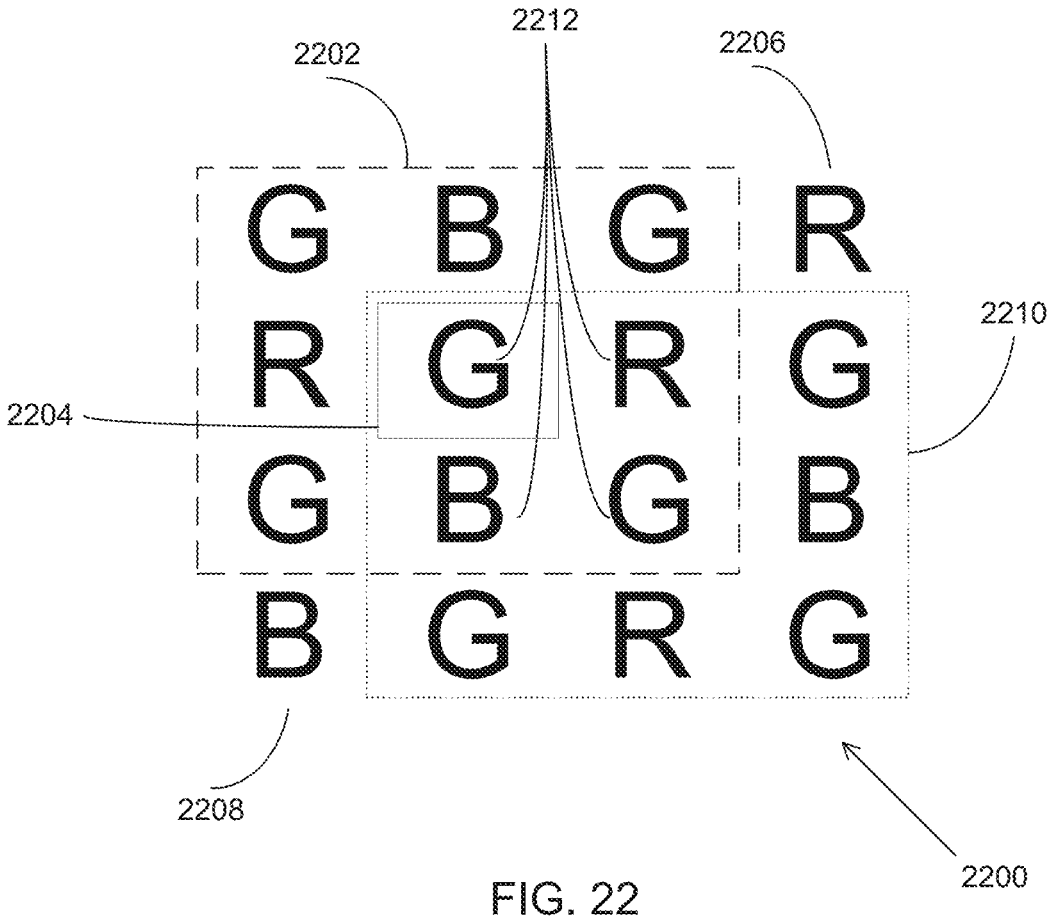


FIG. 22

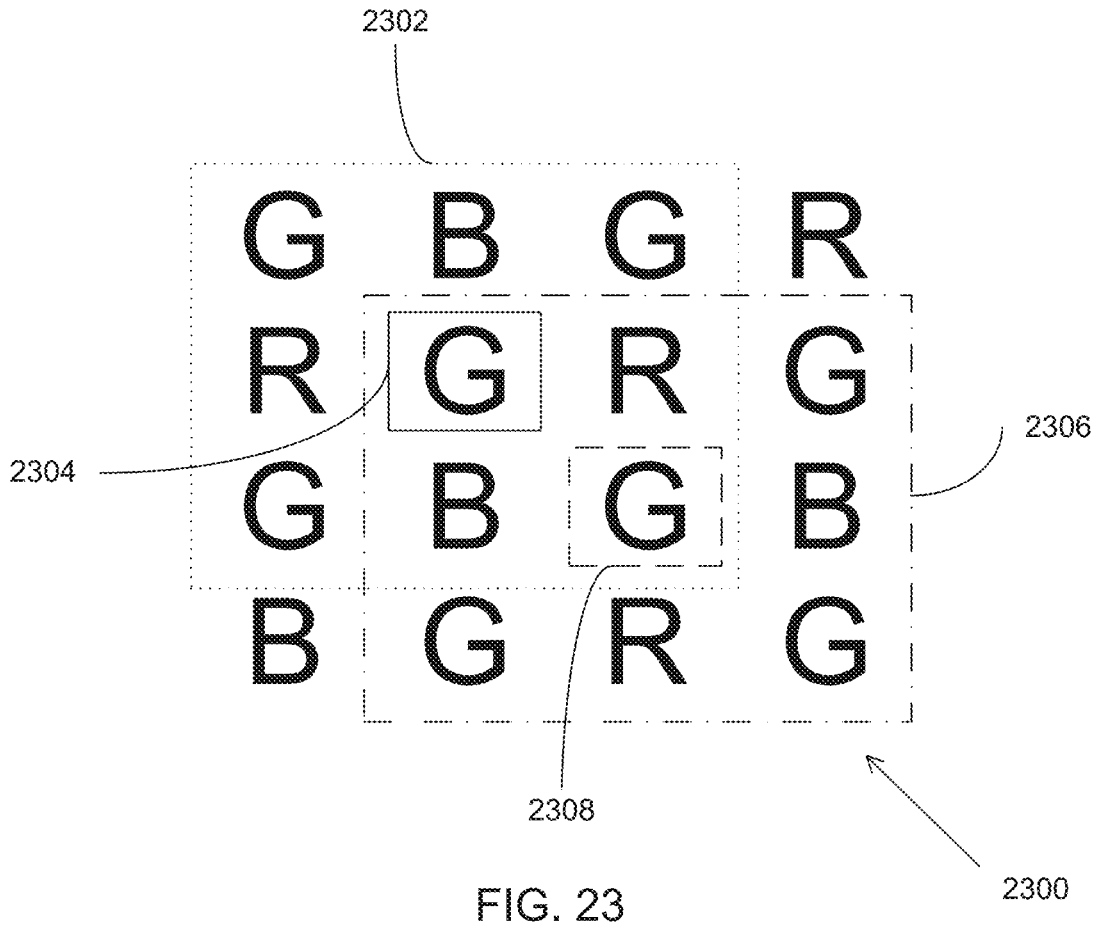


FIG. 23

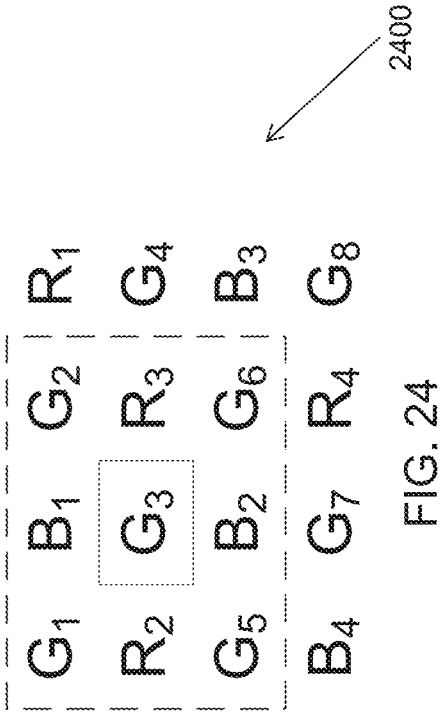


FIG. 24

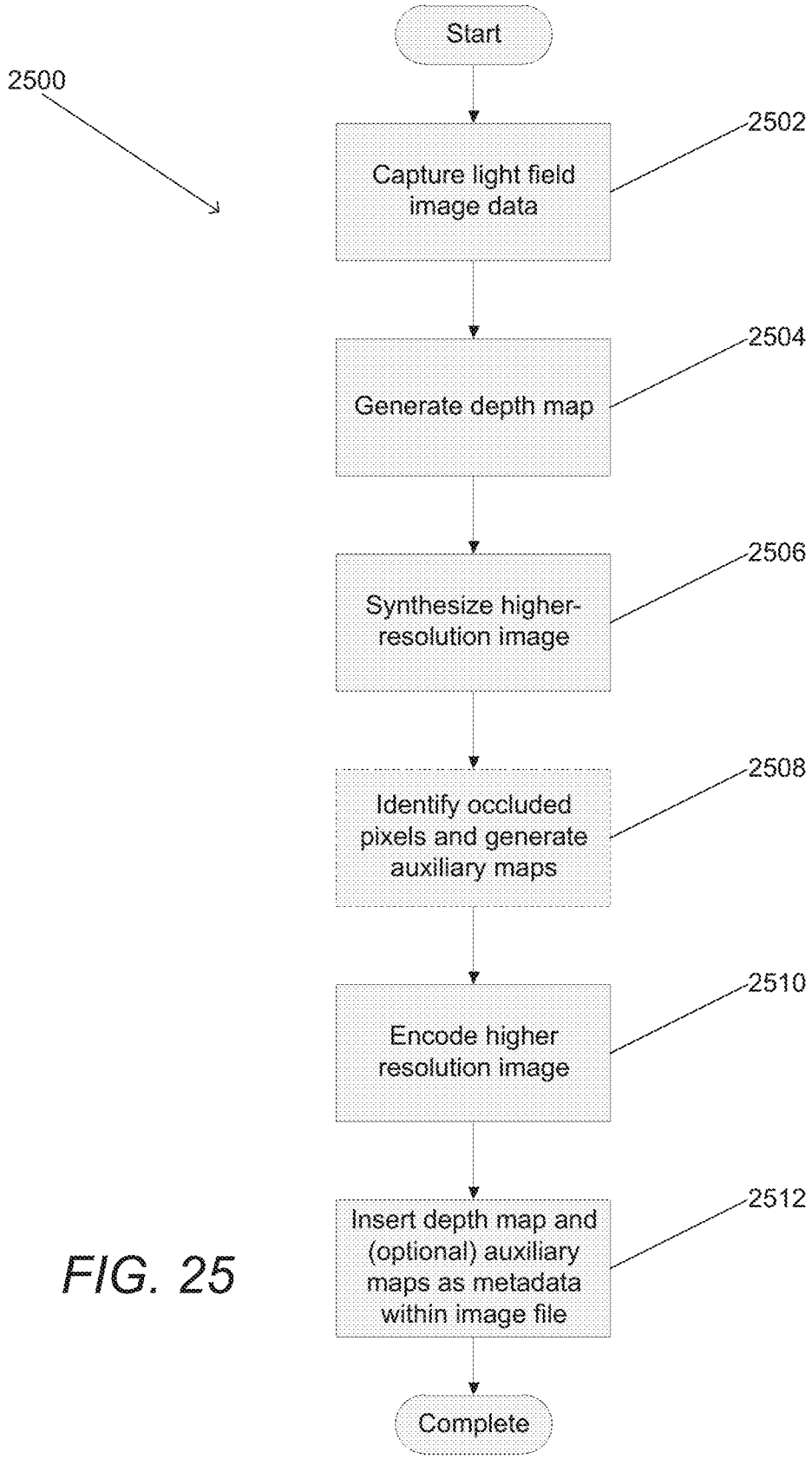

Start
2500
Capture light field image data
2502
Generate depth map
2504
Synthesize higher-resolution image
2506
Identify occluded pixels and generate auxiliary maps
2508
Encode higher resolution image
2510
Insert depth map and (optional) auxiliary maps as metadata within image file
2512
Complete

FIG. 25

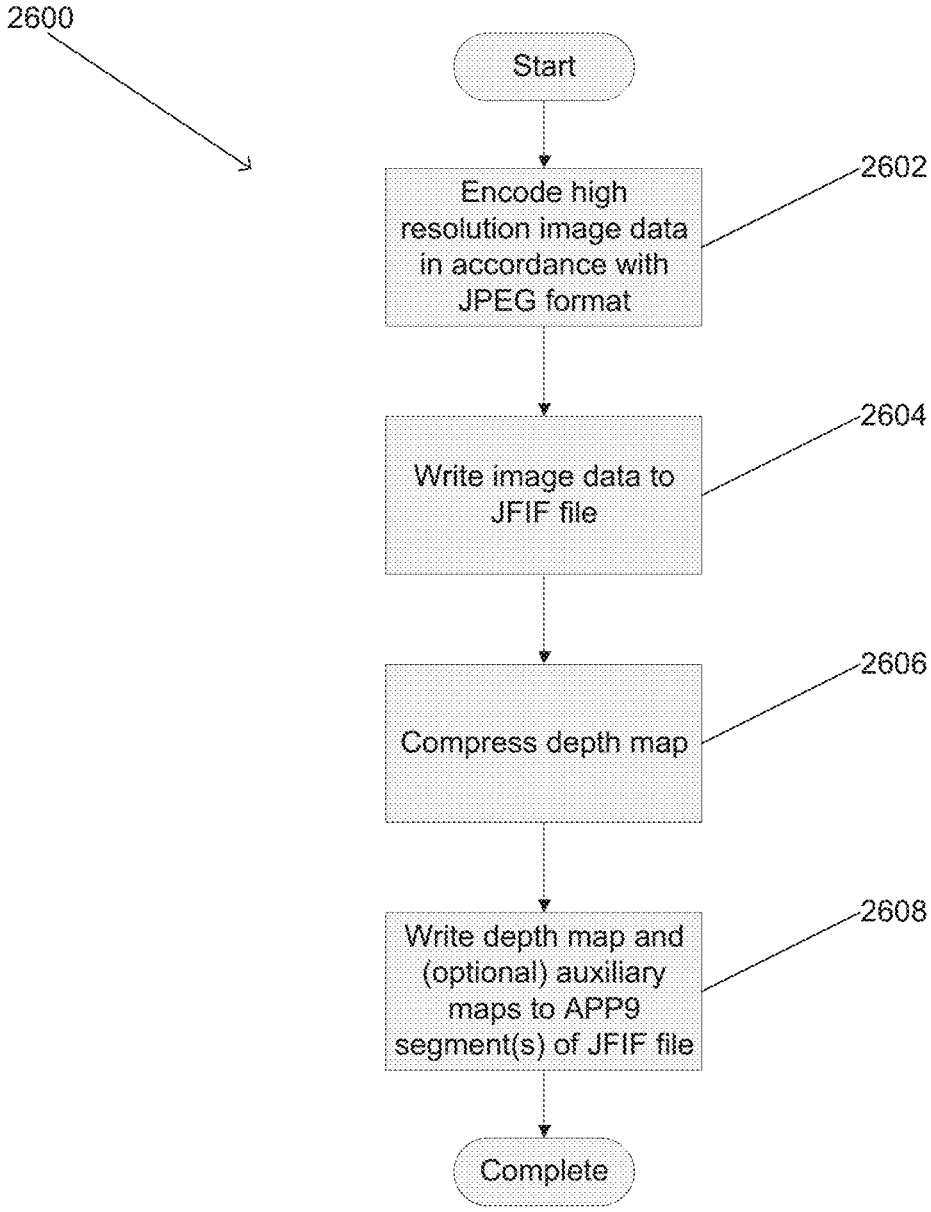

2600
Start
Encode high resolution image data in accordance with JPEG format
2602
Write image data to JFIF file
2604
Compress depth map
2606
Write depth map and (optional) auxiliary maps to APP9 segment(s) of JFIF file
2608
Complete

FIG. 26

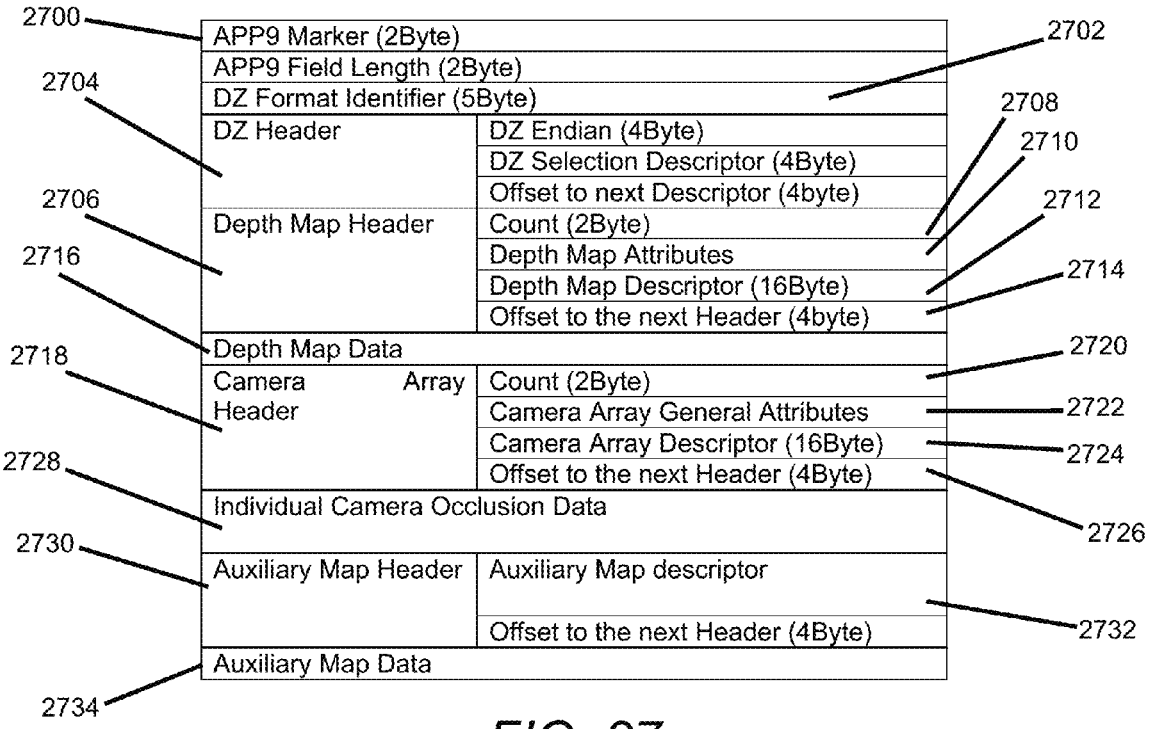


FIG. 27

Name	Byte number	Comments
Reserved for future extensions	Byte 3	
Number of Auxiliary Maps	Byte 2	Describes total number of Auxiliary maps present in the file
Depth Map Camera Array and Auxiliary Maps compression	Byte 1	See detailed description in Figure 6
Depth Map, Camera Array and Auxiliary Maps Selection	Byte 0	See detailed description in Figure 7

FIG. 28

Data Type	Bit number	Selection values
Reserved for Future extensions	Bit 7	
Confidence Map	Bit 7	0: Map is not included 1: Map is Present
Silhouette Edge Map	Bit 5	0: Map is not included 1: Map is Present
Regular Edge Map	Bit 4	0: Map is not included 1: Map is Present
Missing Pixel Map	Bit 3	0: Map is not included 1: Map is Present
Virtual View Point	Bit 2	0: No Virtual View Point 1: Virtual View Point is Present
Camera Array and Per camera Occlusion Data	Bit 1	0: Data is not included 1: Data are Present
Depth Map	Bit 0	0: Data is not included 1: Data is Present

FIG. 29

Reserved for Future extensions	Bit 7	0: Not Compressed 1: Compressed
Confidence Map	Bit 7	0: Not Compressed 1: Compressed
Silhouette Edge Map	Bit 5	0: Not Compressed 1: Compressed
Regular Edge Map	Bit 4	0: Not Compressed 1: Compressed
Missing Pixel Map	Bit 3	0: Not Compressed 1: Compressed
Virtual View Point	Bit 2	0: Not Present 1: Present
Camera Array and Per camera Occlusion Data	Bit 1	0: Not Compressed 1: Compressed
Depth Map	Bit 0	0: Regular JPEG compression 1: Lossless JPEG LS compression

FIG. 30

Attribute name	Attribute Tag ID (4Byte)	Attribute Description
Reserved for future extensions	2-127	
In Focus Plane	1	Define central in focus Depth Plane
F	0	Defines Synthetic Aperture

FIG. 31

Name	Byte Number	Description	Comments
Reserved for future extensions	9-15		
Version	7-8	The most significant byte is used for major revisions, the least significant byte for minor revisions. Version 1.00 is the current revision.	Initial version must support Regular JPEG 8-bit grayscale compression
Identifier	0-6	Zero terminated string "PIDZDH" uniquely identifies this descriptor	

FIG. 32

Name	Byte Numbers	Description	Comments
Compressed Depth Map Data	16 to 16 + Length	Compressed Depth Map Data for Current Segment	
Reserved for future extensions	13-15		
Offset to the next Marker	9-12	Number of bites from current Marker to the next Marker excluding the APP9 marker itself	Next Marker could be either Depth Map Data Segment Marker, Camera Array Marker or Auxiliary Map Marker
Length	7-8	Total APP9 field byte count, including the byte count value (2 bytes), but excluding the APP9 marker itself.	A JPEG marker is limited to 65533 bytes. To represent the whole map multiple depth map markers with the same ID will be following each other
Identifier	0-6	Zero terminated string "PIDZDD" uniquely identifies this descriptor	

FIG. 33

Attribute name	Attribute Tag ID (4Byte)	Bytes	Attribute Description	Comments
Horizontal dimension	0	4	Number of cameras in horizontal direction N	We assume two dimensional Camera Array
Vertical dimension	1	4	Number of cameras in horizontal direction M	
Reference camera position	2	4	Camera number (row major)	
Virtual View Position	3	8	Two floating point numbers Px and Py defining Horizontal and Vertical position of Virtual View point	Constrains: 0<= Px <= N 0<= Py <= M
Number of cameras with occlusion data	4	4	Number of cameras in camera array with occlusion data	

FIG. 34

Name	Byte Number	Description	Comments
Reserved for future extensions	9-15		
Version	7-8	The most significant byte is used for major revisions, the least significant byte for minor revisions. Version 1.00 is the current revision.	Initial version must support Regular JPEG 8-bit grayscale compression
Identifier	0-6	Zero terminated string "PIDZAH" uniquely identifies this descriptor	

FIG. 35

Name	Byte Numbers	Description	Comments
Reserved for future extensions	14-15		
Offset to the next Marker	12-13	Number of bites from current Marker to the next Marker excluding the APP9 marker itself	Next Marker could be either Depth Map Data Segment Marker, Camera Array Marker or Auxiliary Map Marker
Length	10-11	Total APP9 field byte count, including the byte count value (2 bytes), but excluding the APP9 marker itself.	A JPEG marker is limited to 65533 bytes. To represent the whole map multiple depth map markers with the same ID will be following each other
Number of Occluded pixels	8-9	Total Number of Occluded Pixels in this Camera	Currently Camera Array Data is not compressed
Camera Number	7	Camera Number (row major)	
Identifier	0-6	Zero terminated string "PIDZCD" uniquely identifies this descriptor	

FIG. 36

Name	Byte Numbers	Description	Comments
Individual Camera Data for Current Segment	16 to 16 + Length	Occlusion Data	Occlusion Data in current revision is not compressed
Reserved for future extensions	12-15		
Offset to the next Marker	10-11	Number of bites from current Marker to the next Marker excluding the APP9 marker itself	Next Marker could be either Depth Map Data Segment Marker, Camera Array Marker or Auxiliary Map Marker
Length	8-9	Total APP9 field byte count, including the byte count value (2 bytes), but excluding the APP9 marker itself.	A JPEG marker is limited to 65533 bytes. To represent the whole map multiple depth map markers with the same ID will be following each other
Number of Pixels	7- 8	Number of occluded pixels in current segment	
Identifier	0-6	Zero terminated string "PIDZCD" uniquely identifies this descriptor	

FIG. 37

Name	Number of Bytes	Attribute Description	Comments
Depth	1	Depth value	
RGB Color	3	Pixel color	
Pixel coordinates	8	Two integer numbers defining x an y Pixel coordinates	

FIG. 38

Name	Byte Number	Description	Comments
Reserved for future extensions	11-15		
Number of bits per Map	10	Number of bits per pixel for current Auxiliary Map	
Auxiliary Map Type	9	Type of Auxiliary maps	The following types of Auxiliary Maps have been currently defined: 0 = Missing Pixel 1 = Regular Edge 2 = Silhouette Edge 3 = Confidence
Version	7-8	The most significant byte is used for major revisions, the least significant byte for minor revisions. Version 1.00 is the currentrevision.	Initial version must support Regular JPEG 8-bit grayscale compression
Identifier	0-6	Zero terminated string "PIDZAM" uniquely identifies this descriptor	

FIG. 39

Name	Byte Numbers	Description	Comments
Auxiliary Map Data	16 to 16 + Length	Auxiliary data	Not compressed in current revision
Reserved for future extensions	15		
Offset to the next Marker	11-14	Number of bites from current Marker to the next Marker excluding the APP9 marker itself	Next Marker could be either Depth Map Data Segment Marker, Camera Array Marker or Auxiliary Map Marker
Length	9-10	Total APP9 field byte count, including the byte count value (2 bytes), but excluding the APP9 marker itself.	A JPEG marker is limited to 65533 bytes. To represent the whole map multiple depth map markers with the same ID will be following each other
Number of pixels	7-8	Number of pixels in current segment	
Identifier	0-6	Zero terminated string "PIDZAD" uniquely identifies this descriptor	

FIG. 40

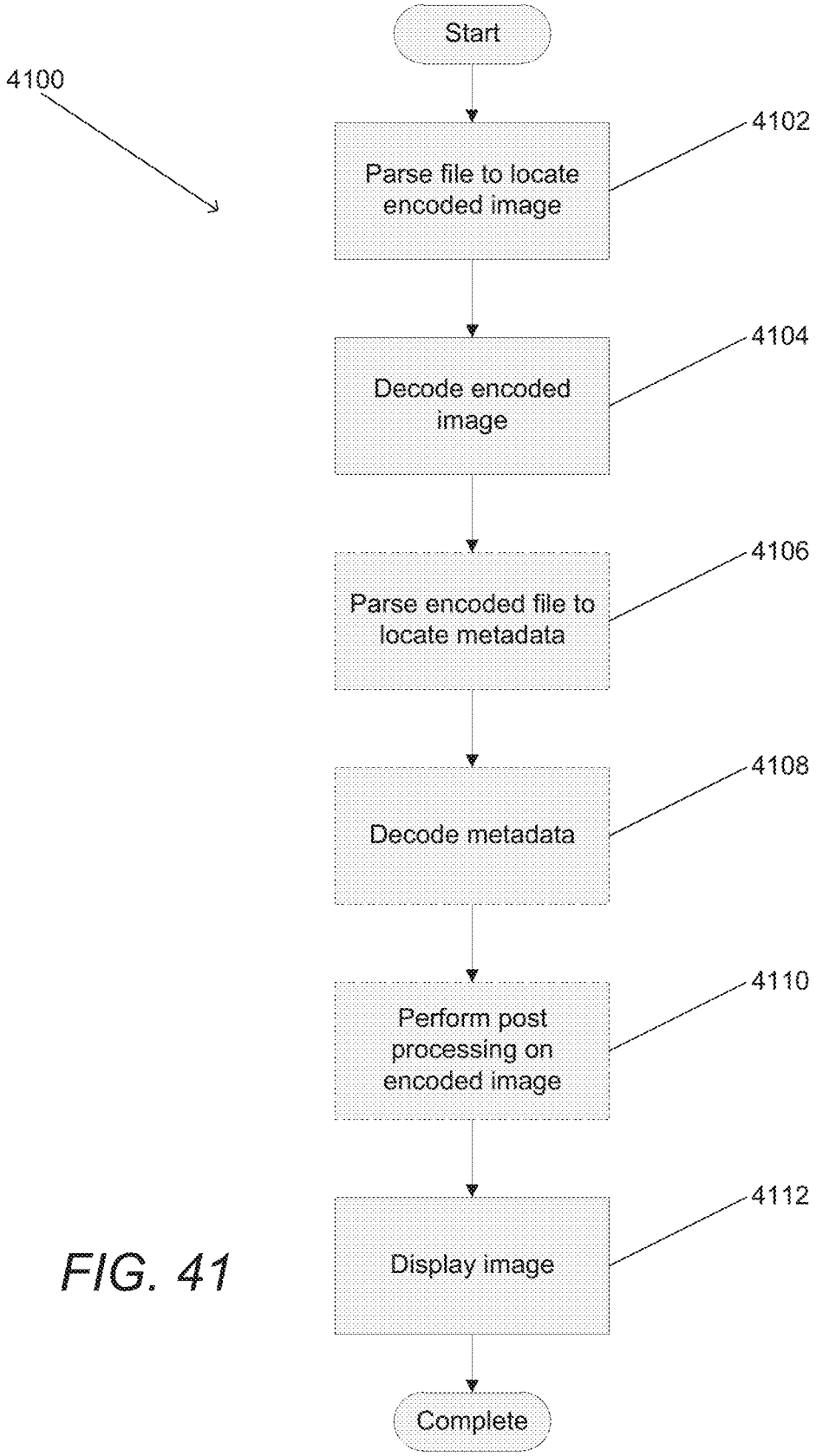

4100
Start
Parse file to locate encoded image
4102
Decode encoded image
4104
Parse encoded file to locate metadata
4106
Decode metadata
4108
Perform post processing on encoded image
4110
Display image
4112
Complete

FIG. 41

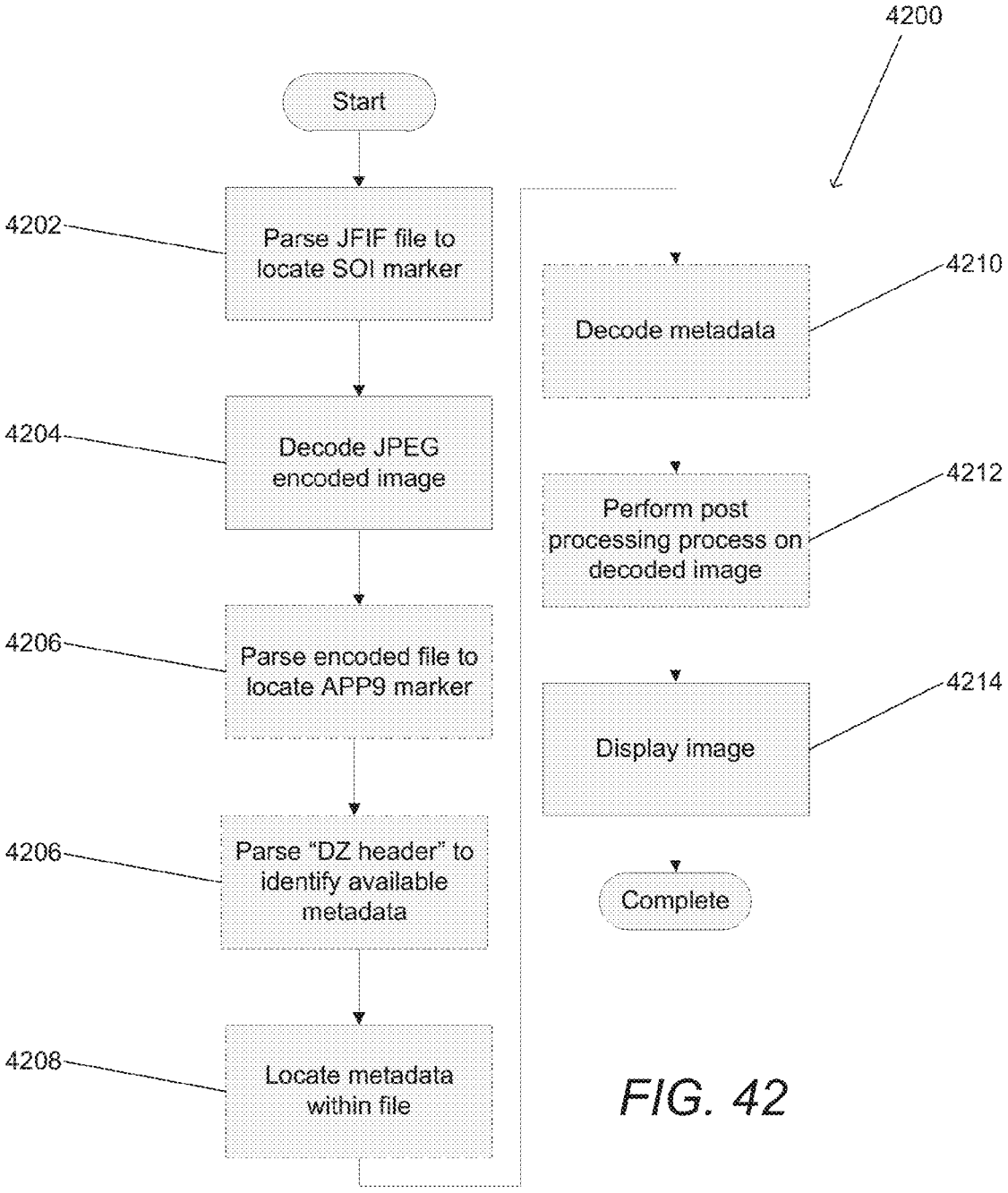

4200
Start
4202
Parse JFIF file to locate SOI marker
4204
Decode JPEG encoded image
4206
Parse encoded file to locate APP9 marker
4206
Parse "DZ header" to identify available metadata
4208
Locate metadata within file
4210
Decode metadata
4212
Perform post processing process on decoded image
4214
Display image
Complete

FIG. 42

SYSTEMS AND METHODS FOR MEASURING DEPTH USING AN ARRAY OF INDEPENDENTLY CONTROLLABLE CAMERAS

CROSS-REFERENCE TO RELATED APPLICATIONS

The current application claims priority as a continuation of U.S. patent application Ser. No. 14/144,458 entitled "Systems and Methods for Performing Depth Estimation using Image Data from Multiple Spectral Channels", filed Dec. 30, 2013, which is a continuation of U.S. patent application Ser. No. 13/972,881 entitled "Systems and Methods for Parallax Detection and Correction in Images Captured Using Array Cameras that Contain Occlusions using Subsets of Images to Perform Depth Estimation", filed Aug. 21, 2013, which claims priority to U.S. Provisional Patent Application Ser. No. 61/691,666 to Venkataraman et al. entitled "Systems and Methods for Parallax Detection and Correction in Images Captured using Array Cameras", filed Aug. 21, 2012 and U.S. Provisional Patent Application Ser. No. 61/780,906 to Venkataraman et al. entitled "Systems and Methods for Parallax Detection and Correction in Images Captured using Array Cameras", filed Mar. 13, 2013. The disclosures of U.S. patent application Ser. Nos. 14/144,458, 13/972,881 and U.S. Provisional Patent Application Ser. Nos. 61/691,666 and 61/780,906 are hereby incorporated by reference herein in their entirety.

FIELD OF THE INVENTION

The present invention generally relates to digital cameras and more specifically to the detection and correction of parallax in images captured using array cameras.

BACKGROUND

Binocular viewing of a scene creates two slightly different images of the scene due to the different fields of view of each eye. These differences, referred to as binocular disparity (or parallax), provide information that can be used to calculate depth in the visual scene, providing a major means of depth perception. The impression of depth associated with stereoscopic depth perception can also be obtained under other conditions, such as when an observer views a scene with only one eye while moving. The observed parallax can be utilized to obtain depth information for objects in the scene. Similar principles in machine vision can be used to gather depth information.

Two or more cameras separated by a distance can take pictures of the same scene and the captured images can be compared by shifting the pixels of two or more images to find parts of the images that match. The amount an object shifts between different camera views is called the disparity, which is inversely proportional to the distance to the object. A disparity search that detects the shift of an object in multiple images can be used to calculate the distance to the object based upon the baseline distance between the cameras and the focal length of the cameras involved. The approach of using two or more cameras to generate stereoscopic three-dimensional images is commonly referred to as multi-view stereo.

Multi-view stereo can generally be described in terms of the following components: matching criterion, aggregation method, and winner selection. The matching criterion is used as a means of measuring the similarity of pixels or regions across different images. A typical error measure is the RGB or intensity difference between images (these differences can be squared, or robust measures can be used). Some methods compute subpixel disparities by computing the analytic minimum of the local error surface or use gradient-based techniques. One method involves taking the minimum difference between a pixel in one image and the interpolated intensity function in the other image. The aggregation method refers to the manner in which the error function over the search space is computed or accumulated. The most direct way is to apply search windows of a fixed size over a prescribed disparity space for multiple cameras. Others use adaptive windows, shiftable windows, or multiple masks. Another set of methods accumulates votes in 3D space, e.g., a space sweep approach and voxel coloring and its variants. Once the initial or aggregated matching costs have been computed, a decision is made as to the correct disparity assignment for each pixel. Local methods do this at each pixel independently, typically by picking the disparity with the minimum aggregated value. Cooperative/competitive algorithms can be used to iteratively decide on the best assignments. Dynamic programming can be used for computing depths associated with edge features or general intensity similarity matches. These approaches can take advantage of one-dimensional ordering constraints along the epipolar line to handle depth discontinuities and unmatched regions. Yet another class of methods formulate stereo matching as a global optimization problem, which can be solved by global methods such as simulated annealing and graph cuts.

More recently, researches have used multiple cameras spanning a wider synthetic aperture to capture light field images (e.g. the Stanford Multi-Camera Array). A light field, which is often defined as a 4D function characterizing the light from all direction at all points in a scene, can be interpreted as a two-dimensional (2D) collection of 2D images of a scene. Due to practical constraints, it is typically difficult to simultaneously capture the collection of 2D images of a scene that form a light field. However, the closer in time at which the image data is captured by each of the cameras, the less likely that variations in light intensity (e.g. the otherwise imperceptible flicker of fluorescent lights) or object motion will result in time dependent variations between the captured images. Processes involving capturing and resampling a light field can be utilized to simulate cameras with large apertures. For example, an array of M×N cameras pointing at a scene can simulate the focusing effects of a lens as large as the array. Use of camera arrays in this way can be referred to as synthetic aperture photography.

While stereo matching was originally formulated as the recovery of 3D shape from a pair of images, a light field captured using a camera array can also be used to reconstruct a 3D shape using similar algorithms to those used in stereo matching. The challenge, as more images are added, is that the prevalence of partially occluded regions (pixels visible in some but not all images) also increases.

SUMMARY OF THE INVENTION

Systems and methods in accordance with embodiments of the invention can perform parallax detection and correction in images captured using array cameras. An embodiment of the method of the invention for estimating distances to objects within a scene from a light field comprising a set of images captured from different viewpoints using a processor configured by an image processing application includes: selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints; normalizing the set of images to increase the similarity of corresponding

pixels within the set of images; and determining initial depth estimates for pixel locations in an image from the reference viewpoint using at least a subset of the set of images, where an initial depth estimate for a given pixel location in the image from the reference viewpoint is determined by: identifying pixels in the at least a subset of the set of images that correspond to the given pixel location in the image from the reference viewpoint based upon expected disparity at a plurality of depths; comparing the similarity of the corresponding pixels identified at each of the plurality of depths; and selecting the depth from the plurality of depths at which the identified corresponding pixels have the highest degree of similarity as an initial depth estimate for the given pixel location in the image from the reference viewpoint. In addition, the method includes identifying corresponding pixels in the set of images using the initial depth estimates; comparing the similarity of the corresponding pixels in the set of images to detect mismatched pixels. When an initial depth estimate does not result in the detection of a mismatch between corresponding pixels in the set of images, selecting the initial depth estimate as the current depth estimate for the pixel location in the image from the reference viewpoint. When an initial depth estimate results in the detection of a mismatch between corresponding pixels in the set of images, selecting the current depth estimate for the pixel location in the image from the reference viewpoint by: determining a set of candidate depth estimates using a plurality of different subsets of the set of images; identifying corresponding pixels in each of the plurality of subsets of the set of images based upon the candidate depth estimates; and selecting the candidate depth of the subset having the most similar corresponding pixels as the current depth estimate for the pixel location in the image from the reference viewpoint.

In a further embodiment, selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints includes selecting a viewpoint from the set consisting of: the viewpoint of one of the images; and a virtual viewpoint.

In another embodiment, a pixel in a given image from the set of images that corresponds to a pixel location in the image from the reference viewpoint is determined by applying a scene dependent shift to the pixel location in the image from the reference viewpoint that is determined based upon: the depth estimate of the pixel location in the image from the reference viewpoint; and the baseline between the viewpoint of the given image and the reference viewpoint.

In a still further embodiment, the subsets of the set of images used to determine the set of candidate depth estimates are selected based upon the viewpoints of the images in the sets of images to exploit patterns of visibility characteristic of natural scenes that are likely to result in at least one subset in which a given pixel location in the image from the reference viewpoint is visible in each image in the subset.

In still another embodiment, the set of images are captured within multiple color channels; selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints includes selecting one of the images as a reference image and selecting the viewpoint of the reference image as the reference viewpoint; and the subsets of the set of images used to determine the set of candidate depth estimates are selected so that the same number of images in the color channel containing the reference image appears in each subset.

In a yet further embodiment, the subsets of the set of images used to determine the set of candidate depth estimates are also selected so that there are at least two images in the color channels that do not contain the reference image in each subset.

Yet another embodiment also includes determining the visibility of the pixels in the set of images from the reference viewpoint by: identifying corresponding pixels in the set of images using the current depth estimates; and determining that a pixel in a given image is not visible in the image from the reference viewpoint when the pixel fails a photometric similarity criterion determined based upon a comparison of corresponding pixels.

In a further embodiment again, selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints includes selecting one of the images in the set of images as a reference image and selecting the viewpoint of the reference image as the reference viewpoint; and determining that a pixel in a given image is not visible in the image from the reference viewpoint when the pixel fails a photometric similarity criterion determined based upon a comparison of corresponding pixels further includes comparing the pixel in the given image to the corresponding pixel in the reference image.

In another embodiment again, the photometric similarity criterion includes a similarity threshold that adapts based upon at least the intensity of at least one of the pixel in the given image and the pixel in the reference image.

In a further additional embodiment, the photometric similarity criterion includes a similarity threshold that adapts as a function of the photometric distance between the corresponding pixel from the reference image and the corresponding pixel that is most similar to the pixel from the reference image.

In another additional embodiment, the photometric similarity criterion includes a similarity threshold that adapts based upon the signal to noise ratio of the pixel in the reference image.

In a still yet further embodiment, adapting the similarity threshold based upon the signal to noise ratio is approximated by scaling the photometric distance of the corresponding pixel from the reference image and the corresponding pixel that is most similar to the pixel from the reference image is and applying an offset to obtain an appropriate threshold.

In still yet another embodiment, the set of images includes images captured in a plurality of color channels and the reference image is an image captured in a first color channel and the given image is in the second color channel; determining that a pixel in a given image is not visible in the reference viewpoint when the pixel fails a photometric similarity criterion determined based upon a comparison of corresponding pixels further includes: selecting an image in the second color channel in which the corresponding pixel in the image from the reference viewpoint is visible as a reference image for the second color channel; and comparing the pixel in the given image to the corresponding pixel in the reference image for the second color channel.

In a still further embodiment again, selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints includes selecting a virtual viewpoint as the reference viewpoint; and determining that a pixel in a given image is not visible in the image from the reference viewpoint when the pixel fails a photometric similarity criterion determined based upon a comparison of corresponding pixels further includes: selecting an image adjacent the virtual viewpoint as a reference image; and comparing the pixel in the given image to the corresponding pixel in the reference image.

In still another embodiment again, the image adjacent the virtual viewpoint is selected based upon the corresponding pixel in the selected image to the pixel from the given image being visible in an image from the reference viewpoint.

A yet further embodiment again also includes updating the depth estimate for a given pixel location in the image from the reference viewpoint based upon the visibility of the pixels in the set of images from the reference viewpoint by: generating an updated subset of the set of images using images in which the given pixel location in the image from the reference viewpoint is determined to be visible based upon the current depth estimate for the given pixel; identifying pixels in the updated subset of the set of images that correspond to the given pixel location in the image from the reference viewpoint based upon expected disparity at a plurality of depths; comparing the similarity of the corresponding pixels in the updated subset of images identified at each of the plurality of depths; and selecting the depth from the plurality of depths at which the identified corresponding pixels in the updated subset of the set of images have the highest degree of similarity as an updated depth estimate for the given pixel location in the image from the reference viewpoint.

In yet another embodiment again, the subsets of the set of images are pairs of images; and the updated subset of the set of images includes at least three images.

In a still further additional embodiment, normalizing the set of images to increase the similarity of corresponding pixels within the set of images further includes utilizing calibration information to correct for photometric variations and scene-independent geometric distortions in the images in the set of images, and rectification of the images in the set of images

In still another additional embodiment, normalizing the set of images to increase the similarity of corresponding pixels within the set of images further includes resampling the images to increase the similarity of corresponding pixels in the set of images; and the scene-independent geometric corrections applied to the images are determined at a sub-pixel resolution.

In a yet further additional embodiment, utilizing calibration information to correct for photometric variations further includes performing any one of the normalization processes selected from the group consisting of: Black Level calculation and adjustment; vignetting correction; lateral color correction; and temperature normalization.

In yet another additional embodiment, the scene-independent geometric corrections also include rectification to account for distortion and rotation of lenses in an array of cameras that captured the set of images.

In a further additional embodiment again, a cost function is utilized to determine the similarity of corresponding pixels.

In another additional embodiment again, determining the similarity of corresponding pixels further includes spatially filtering the calculated costs.

In another further embodiment, selecting the depth from the plurality of depths at which the identified corresponding pixels have the highest degree of similarity as an initial depth estimate for the given pixel location in the image from the reference viewpoint further includes selecting the depth from the plurality of depths at which the filtered cost function for the identified corresponding pixels indicates the highest level of similarity.

In still another further embodiment, the cost function utilizes at least one similarity measure selected from the group consisting of: the L1 norm of a pair of corresponding pixels; the L2 norm of a pair of corresponding pixels; and the variance of a set of corresponding pixels.

In yet another further embodiment, the set of images are captured within multiple color channels and the cost function determines the similarity of pixels in each of the multiple color channels.

Another further embodiment again also includes generating confidence metrics for the current depth estimates for pixel locations in the image from the reference viewpoint.

In another further additional embodiment, the confidence metric encodes a plurality of confidence factors.

Still yet another further embodiment also includes filtering the depth map based upon the confidence map.

Still another further embodiment again also includes detecting occlusion of pixels in images within the set of images that correspond to specific pixel locations in the image from the reference viewpoint based upon the initial depth estimates by searching along lines parallel to the baselines between the reference viewpoint and the viewpoints of the images in the set of images to locate occluding pixels; when an initial depth estimate results in the detection of a corresponding pixel in at least one image being occluded, selecting the current depth estimate for the pixel location in the image from the reference viewpoint by: determining a set of candidate depth estimates using a plurality of different subsets of the set of images that exclude the at least one image in which the given pixel is occluded; identifying corresponding pixels in each of the plurality of subsets of the set of images based upon the candidate depth estimates; and selecting the candidate depth of the subset having the most similar corresponding pixels as the current depth estimate for the pixel location in the image from the reference viewpoint.

In still another further additional embodiment, searching along lines parallel to the baselines between the reference viewpoint and the viewpoints of the images in the set of images to locate occluding pixels further includes determining that a pixel corresponding to a pixel location (x_1, y_1) in an image from the reference viewpoint is occluded in an alternate view image by a pixel location (x_2, y_2) in the image from the reference viewpoint when

$$|s_2-s_1-\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}|\leq \text{threshold}$$

where s_1 and s_2 are scene dependent geometric shifts applied to pixel locations (x_1, y_1) and pixel (x_2, y_2) to shift the pixels along a line parallel to the baseline between the reference viewpoint and the viewpoint of the alternate view image to shift the pixels into the viewpoint of the alternate view image based upon the initial depth estimates for each pixel.

In yet another further embodiment again, the decision to designate a pixel as being occluded considers at least one of the similarity of the pixels and the confidence of the estimated depths of the pixels (x_1, y_1) and (x_2, y_2).

In a specific embodiment, a cost function is utilized to determine the similarity of corresponding pixels.

In another specific embodiment, determining the similarity of corresponding pixels further comprises spatially filtering the calculated costs.

In a further specific embodiment, the spatial filtering of the calculated costs utilizes a filter selected from the group consisting of: a fixed-coefficient filter; and an edge-preserving filter.

In a still further specific embodiment, selecting the depth from the plurality of depths at which the identified corresponding pixels have the highest degree of similarity as an initial depth estimate for the given pixel location in the image from the reference viewpoint further includes selecting the depth from the plurality of depths at which the filtered cost function for the identified corresponding pixels indicates the highest level of similarity.

In still another specific embodiment, the set of images are captured within a single color channel and the cost function is a function of the variance of the corresponding pixel.

In a yet further specific embodiment, the cost function is an aggregated cost function CV(x, y, d) over each image i in the set of images that includes the following term

$$CV(x, y, d) = \sum_i \frac{Cost^{i,Ref}(x, y, d) \times V^{i,Ref}(x, y)}{\text{number of visible cameras at } (x, y)}$$

where $Cost^{i,Ref}(x, y, d)$ is a similarity measure (i.e. the cost function),

d is depth of pixel (x, y), and

$V^{i,Ref}(x\ y)$ is the visibility of pixel (x, y) and initially $V^{i,Ref}(x, y)=1$ for all cameras.

In a further specific embodiment again, the individual costs $Cost^{i,Ref}$ (x, y, d) are computed based on each disparity hypothesis d for each pixel (x, y) for cameras i, Ref as follows:

$$Cost^{i,Ref}(x,y,d)=S\{I^i(x,y,d),I^{Ref}(x,y,d)\}$$

where S is the similarity measure (for example), and

I^i is the calibrated image i after geometric calibration.

In yet another specific embodiment, the aggregated cost considers the similarity of the shifted images at the candidate depth as follows:

$$CV(x, y, d) = \sum_{k \in K} \frac{(x, y) Cost^{k,Ref}(x, y, d) \times V^{k,Ref}}{\text{number of cameras in } K} + \sum_{i,j \in L} \frac{Cost^{i,j}(x, y, d) \times V^{i,Ref}(x, y) \times V^{j,Ref}(x, y)}{\text{number of pairs of cameras in } L}$$

where K is a set of cameras in the same spectral channel as the reference camera,

L is a set of pairs of cameras, where both cameras in each pair are in the same spectral channel (which can be a different spectral channel to the reference camera where the light field includes image data in multiple spectral channels),

$Cost^{k,Ref}$ (x, y, d)=S{ImageRef(x, y), ShiftedImage^k(x, y, d)}, and

$Cost^{i,j}$(x, y, d)=S{ShiftedImage^i(x, y, d), ShiftedImage^j(x, y, d)}

In a further specific embodiment again, the aggregated cost function is spatially filtered using a filter so that the weighted aggregated cost function is as follows:

$$FilteredCV(x, y, d) = \frac{1}{Norm} \sum_{\substack{(x_1,y_1) \\ \in N(x,y)}} CV(x_1, y_1, d) \times wd(x, y, x_1, y_1) \times wr(I_{Ref}(x, y) - I_{Ref}(x_1, y_1))$$

where N(x, y) is the immediate neighborhood of the pixel (x, y), which can be square, circular, rectangular, or any other shape appropriate to the requirements of a specific application,

Norm is a normalization term,

I_{Ref}(X, y) is the image data from the reference camera,

wd is a weighting function based on pixel distance, and

wr is a weighting function based on intensity difference.

In a further embodiment, the filter is a box filter and wd and wr are constant coefficients.

In another embodiment, the filter is a bilateral filter and wd and wr are both Gaussian weighting functions.

In a still further embodiment, a depth estimate for a pixel location (x, y) in the image from the reference viewpoint is determined by selecting the depth that minimizes the filtered cost at each pixel location in the depth map as follows:

$$D(x,y)=\text{argmin}\{\text{Filtered}CV(x,y,d)\}$$

In still another embodiment, the set of images are captured within multiple color channels and the cost function incorporates the L1 norm of image data from the multiple color channels.

In a yet further embodiment, the set of images are captured within multiple color channels and the cost function incorporates the L2 norm of image data from the multiple color channels.

In yet another embodiment, the set of images are captured within multiple color channels including at least Red, Green and Blue color channels; selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints comprises selecting one of the images in the Green color channel as a Green reference image and selecting the viewpoint of the Green reference image as the reference viewpoint; and the cost function Cost (x, y, d) for a pixel location (x, y) in the image from the reference viewpoint at a depth d is:

$$\text{Cost}(x,y,d)=\gamma_G(x,y)\cdot\text{Cost}_G(x,y,d)+\gamma_R(x,y)\cdot\text{Cost}_R(x,y,d)+\gamma_B(x,y)\cdot\text{Cost}_B(x,y,d)$$

where $Cost_G$ (x, y, d) is the measure of the similarity of a pixel location (x, y) in the image from the reference viewpoint to corresponding pixels in locations within a set of Green images based upon the depth d,

$Cost_R$ (x, y, d) is the measure of the similarity of corresponding pixels in locations within a set of Red images determined based upon the depth d and the pixel location (x, y) in the image from the reference viewpoint,

$Cost_B$ (x, y, d) is the measure of the similarity of corresponding pixels in locations within a set of Blue images determined based upon the depth d and the pixel location (x, y) in the image from the reference viewpoint, and

γ_G, γ_R, and γ_B are weighting factors for the Green, Red and Blue cost functions respectively.

In a further embodiment again, the $Cost_G$ (x,y,d) uses a similarity measure selected from the group consisting of an L1 norm, an L2 norm, and variance across the pixels in the images in the set of images that are within the Green color channel.

In another embodiment again, the cost measures for the Red ($Cost_R$ (x, y, d)) and Blue color channels ($Cost_B$ (x,y,d)) are determined by calculating the aggregated difference between unique pairs of corresponding pixels in images within the color channel.

In a further additional embodiment, calculating the aggregated difference between each unique pair of corresponding pixels in images within a color channel comprises determining a combination cost metric for unique pairs of corresponding pixels in images within the color channel.

In another additional embodiment, the combination cost metric ($Cost_C$ (x,y,d)) for a Red color channel including four images (C_A, C_B, C_C, and C_D) can be determined as follows:

$$Cost_C(x, y, d) = |C_A(x_A, y_A) - C_B(x_B, y_B)| + |C_A(x_A, y_A) - C_C(x_C, y_C)| + |C_A(x_A, y_A) - C_D(x_D, y_D)| + |C_B(x_B, y_B) - C_C(x_C, y_C)| + |C_B(x_B, y_B) - C_D(x_D, y_D)| + |C_C(x_C, y_C) - C_D(x_D, y_D)|$$

where (x_A,y_A), (x_B,y_B), (x_C,y_C), and (x_D,x_D) are corresponding pixel locations determined based upon the disparity in each of the images C_A, C_B, C_C, and C_D respectively at depth d.

In a still yet further embodiment, the combination cost metric is determined utilizing at least one selected from the group consisting of: the L1 norm of the pixel brightness values; the L2 norm of the pixel brightness values; and the variance in the pixel brightness values.

In still yet another embodiment, the weighting factors γ_C, γ_B, and γ_B are fixed.

In a still further embodiment again, the weighting factors γ_G, γ_R, and γ_B vary spatially with the pixel location (x, y) in the image from the reference viewpoint.

In still another embodiment again, the weighting factors γ_G, γ_R, and γ_B vary based upon the estimated SNR at the pixel location (x, y) in the image from the reference viewpoint; and strong SNR at the pixel location (x, y) in the image from the reference viewpoint is used to reduce the weighting applied to the Red and Blue color channels.

In a further embodiment, the confidence metric encodes a plurality of confidence factors.

In another embodiment, the confidence metric for the depth estimate for a given pixel location in the image from the reference viewpoint comprises at least one confidence factor selected from the group consisting of: an indication that the given pixel is within a textureless region within an image; a measure of the signal to noise ration (SNR) in a region surrounding a given pixel; the number of corresponding pixels used to generate the depth estimate; an indication of the number of depths searched to generate the depth estimate; an indication that the given pixel is adjacent a high contrast edge; and an indication that the given pixel is adjacent a high contrast boundary.

In a still further embodiment, the confidence metric for the depth estimate for a given pixel location in the image from the reference viewpoint comprises at least one confidence factor selected from the group consisting of: an indication that the given pixel lies on a gradient edge; an indication that the corresponding pixels to the given pixel are mismatched; an indication that corresponding pixels to the given pixel are occluded; an indication that depth estimates generated using different reference cameras exceed a threshold for the given pixel; an indication that the depth estimates generated using different subsets of cameras exceed a threshold for the given pixel; an indication as to whether the depth of the given threshold exceeds a threshold; an indication that the given pixel is defective; and an indication that corresponding pixels to the given pixel are defective.

In still another embodiment, the confidence metric for the depth estimate for a given pixel location in the image from the reference viewpoint comprises at least: a measure of the SNR in a region surrounding a given pixel; and the number of corresponding pixels used to generate the depth estimate.

In a yet further embodiment, the confidence metric encodes at least one binary confidence factor.

In yet another embodiment, the confidence metric encodes at least one confidence factor represented as a range of degrees of confidence.

In a further embodiment again, the confidence metric encodes at least one confidence factor determined by comparing the similarity of the pixels in the set of images that were used to generate the finalized depth estimate for a given pixel location in the image from the reference viewpoint.

In another embodiment again, a cost function is utilized to generate a cost metric indicating the similarity of corresponding pixels; and comparing the similarity of the pixels in the set of images that were used to generate the depth estimate for a given pixel location in the image from the reference viewpoint further comprises: applying a threshold to a cost metric of the pixels in the set of images that were used to generate the finalized depth estimate for a given pixel location in the image from the reference viewpoint; and when the cost metric exceeds the threshold, assigning a confidence metric that indicates that the finalized depth estimate for the given pixel location in the image from the reference viewpoint was generated using at least one pixel in the set of images that is a problem pixel.

In a further additional embodiment, the threshold is modified based upon at least one of: a mean intensity of a region surrounding the given pixel location in the image from the reference viewpoint; and noise statistics for at least one sensor used to capture the set of images.

In a still yet further embodiment, the mean intensity of a region surrounding the given pixel location in the image from the reference viewpoint is calculated using a spatial box N×N averaging filter centered around the given pixel.

In still yet another embodiment, the set of images are captured within multiple color channels including at least Red, Green and Blue color channels; selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints comprises selecting one of the images in the Green color channel as a Green reference image and selecting the viewpoint of the Green reference image as the reference viewpoint; and the mean intensity is used to determine the noise statistics for the Green channel using a table that relates a particular mean at a particular exposure and gain to a desired threshold.

In a still further embodiment again, selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints comprises selecting one of the images as a reference image and selecting the viewpoint of the reference image as the reference viewpoint; and a cost function is utilized to generate a cost metric indicating the similarity of corresponding pixels; a confidence metric based upon general mismatch is obtained using the following formula:

$$\text{Confidence}(x,y)=F(\text{Cost}_{min}(x,y),\text{Cost}^{d}(x,y),I(x,y)^{cam},\text{Sensor})$$

where $\text{Cost}_{min}(x, y)$ is the minimum cost of a disparity search over the desired depth range,

- $\text{Cost}^{d}(x, y)$ denotes that cost data from any depth or depths (beside the minimum depth),
- $I(x, y)^{cam}$ image data captured by any camera can be utilized to augment the confidence;
- Sensor is the sensor prior, which can include known properties of the sensor, such as (but not limited to) noise statistics or characterization, defective pixels, properties of the sensor affecting any captured images (such as gain or exposure),
- Camera intrinsics is the camera intrinsic, which specifies elements intrinsic to the camera and camera array that can impact confidence including (but not limited to) the baseline separation between cameras in the array (affects precision of depth measurements), and the arrange-

ment of the color filters (affects performance in the occlusion zones in certain scenarios).

In still another embodiment again, selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints comprises selecting one of the images as a reference image and selecting the viewpoint of the reference image as the reference viewpoint; and a cost function is utilized to generate a cost metric indicating the similarity of corresponding pixels; and a confidence metric based upon general mismatch is obtained using the following formula:

$$\text{Confidence}(x, y) = a \times \frac{Cost_{\min}(x, y)}{Avg(x, y)} + \text{offset}$$

where Avg(x,y) is the mean intensity of the reference image in a spatial neighborhood surrounding (x, y), or an estimate of the mean intensity in the neighborhood, that is used to adjust the confidence based upon the intensity of the reference image in the region of (x, y),

a and offset are empirically chosen scale and offset factors used to adjust the confidence with prior information about the gain and noise statistics of the sensor.

a and offset are empirically chosen scale and offset factors used to adjust the confidence with prior information about the gain and noise statistics of at least one sensor used to capture images in the set of images.

In a yet further embodiment again, generating confidence metrics for the depth estimates for pixel locations in the image from the reference viewpoint includes determining at least one sensor gain used to capture at least one of the set of images and adjusting the confidence metrics based upon the sensor gain.

In yet another embodiment again, generating confidence metrics for the depth estimates for pixel locations in the image from the reference viewpoint comprises determining at least one exposure time used to capture at least one of the set of images and adjusting the confidence metrics based upon the sensor gain.

A still further additional embodiment also includes outputting a depth map containing the finalized depth estimates for pixel locations in the image from the reference viewpoint, and outputting a confidence map containing confidence metrics for the finalized depth estimates contained within the depth map.

Still another additional embodiment also includes filtering the depth map based upon the confidence map.

Yet another further additional embodiment includes estimating distances to objects within a scene from the light field comprising a set of images captured from different viewpoints using a processor configured by an image processing application by: selecting a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints; normalizing the set of images to increase the similarity of corresponding pixels within the set of images; determining initial depth estimates for pixel locations in an image from the reference viewpoint using at least a subset of the set of images, where an initial depth estimate for a given pixel location in the image from the reference viewpoint is determined by: identifying pixels in the at least a subset of the set of images that correspond to the given pixel location in the image from the reference viewpoint based upon expected disparity at a plurality of depths; comparing the similarity of the corresponding pixels identified at each of the plurality of depths; and selecting the depth from the plurality of depths at which the identified corresponding pixels have the highest degree of similarity as an initial depth estimate for the given pixel location in the image from the reference viewpoint. In addition, the process of estimating distances further includes identifying corresponding pixels in the set of images using the initial depth estimates; comparing the similarity of the corresponding pixels in the set of images to detect mismatched pixels; when an initial depth estimate does not result in the detection of a mismatch between corresponding pixels in the set of images, selecting the initial depth estimate as the current depth estimate for the pixel location in the image from the reference viewpoint; and when an initial depth estimate results in the detection of a mismatch between corresponding pixels in the set of images, selecting the current depth estimate for the pixel location in the image from the reference viewpoint by: determining a set of candidate depth estimates using a plurality of different subsets of the set of images; identifying corresponding pixels in each of the plurality of subsets of the set of images based upon the candidate depth estimates; and selecting the candidate depth of the subset having the most similar corresponding pixels as the current depth estimate for the pixel location in the image from the reference viewpoint. The process further including determining the visibility of the pixels in the set of images from the reference viewpoint by: identifying corresponding pixels in the set of images using the current depth estimates; and determining that a pixel in a given image is not visible in the reference viewpoint when the pixel fails a photometric similarity criterion determined based upon a comparison of corresponding pixels; and fusing pixels from the set of images using the processor configured by the image processing application based upon the depth estimates to create a fused image having a resolution that is greater than the resolutions of the images in the set of images by: identifying the pixels from the set of images that are visible in an image from the reference viewpoint using the visibility information; and applying scene dependent geometric shifts to the pixels from the set of images that are visible in an image from the reference viewpoint to shift the pixels into the reference viewpoint, where the scene dependent geometric shifts are determined using the current depth estimates; and fusing the shifted pixels from the set of images to create a fused image from the reference viewpoint having a resolution that is greater than the resolutions of the images in the set of images.

Another further embodiment also includes synthesizing an image from the reference viewpoint using the processor configured by the image processing application to perform a super resolution process based upon the fused image from the reference viewpoint, the set of images captured from different viewpoints, the current depth estimates, and the visibility information.

A further embodiment of the invention includes a processor, and memory containing a set of images captured from different viewpoints and an image processing application. In addition, the image processing application configures the processor to: select a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints; normalize the set of images to increase the similarity of corresponding pixels within the set of images; determine initial depth estimates for pixel locations in an image from the reference viewpoint using at least a subset of the set of images, where an initial depth estimate for a given pixel location in the image from the reference viewpoint is determined by: identifying pixels in the at least a subset of the set of images that correspond to the given pixel location in the image from the reference viewpoint based upon expected disparity at a plurality of depths; comparing the similarity of

the corresponding pixels identified at each of the plurality of depths; and selecting the depth from the plurality of depths at which the identified corresponding pixels have the highest degree of similarity as an initial depth estimate for the given pixel location in the image from the reference viewpoint. The application further configures the processor to identify corresponding pixels in the set of images using the initial depth estimates; compare the similarity of the corresponding pixels in the set of images to detect mismatched pixels. When an initial depth estimate does not result in the detection of a mismatch between corresponding pixels in the set of images, the application configures the processor to select the initial depth estimate as the current depth estimate for the pixel location in the image from the reference viewpoint. When an initial depth estimate results in the detection of a mismatch between corresponding pixels in the set of images, the application configures the processor to select the current depth estimate for the pixel location in the image from the reference viewpoint by: determining a set of candidate depth estimates using a plurality of different subsets of the set of images; identifying corresponding pixels in each of the plurality of subsets of the set of images based upon the candidate depth estimates; and selecting the candidate depth of the subset having the most similar corresponding pixels as the current depth estimate for the pixel location in the image from the reference viewpoint.

In another embodiment, the image processing application further configures the processor to: determine the visibility of the pixels in the set of images from the reference viewpoint by: identifying corresponding pixels in the set of images using the current depth estimates; and determining that a pixel in a given image is not visible in the reference viewpoint when the pixel fails a photometric similarity criterion determined based upon a comparison of corresponding pixels; and fuse pixels from the set of images using the depth estimates to create a fused image having a resolution that is greater than the resolutions of the images in the set of images by: identifying the pixels from the set of images that are visible in an image from the reference viewpoint using the visibility information; and applying scene dependent geometric shifts to the pixels from the set of images that are visible in an image from the reference viewpoint to shift the pixels into the reference viewpoint, where the scene dependent geometric shifts are determined using the current depth estimates; and fusing the shifted pixels from the set of images to create a fused image from the reference viewpoint having a resolution that is greater than the resolutions of the images in the set of images.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** conceptual illustrates of an array camera in accordance with an embodiment of the invention.

FIG. **1**A conceptually illustrates an array camera module in accordance with an embodiment of the invention.

FIG. **1**C conceptually illustrates a color filter pattern for a 4×4 array camera module in accordance with an embodiment of the invention.

FIG. **2** conceptually illustrates capturing image data using a reference camera and an alternate view camera.

FIGS. **3**A and **3**B conceptually illustrate the effect of parallax in images of a scene captured by a reference camera and an alternate view camera.

FIG. **4** is a flowchart illustrating a process for generating a depth map from a captured light field including a plurality of images captured from different viewpoints in accordance with an embodiment of the invention.

FIG. **5** is a flowchart of a process for normalizing captured image data in accordance with an embodiment of the invention.

FIG. **6** is a flowchart of a process for iteratively refining a depth map based upon visibility information in accordance with embodiments of the invention.

FIG. **7** conceptually illustrates a subset of cameras within an array camera that can be utilized to generate estimates of distances to objects within a scene in accordance with an embodiment of the invention.

FIG. **8** is a flowchart illustrating a process for performing a disparity search using visibility information in accordance with an embodiment of the invention.

FIG. **8**A is a flowchart illustrating a process for estimating depth using images captured by subsets of cameras in a camera array in accordance with an embodiment of the invention.

FIGS. **8**B-**8**I conceptually illustrate subsets of cameras in a 5×5 array camera that can be utilized to obtain depth estimates in accordance with embodiments of the invention.

FIGS. **8**J-**8**M conceptually illustrate subsets of cameras in a 4×4 array camera that can be utilized to obtain depth estimates in accordance with embodiments of the invention.

FIG. **9** conceptually illustrates a process for searching an epipolar line for pixels that occlude a given pixel in accordance with an embodiment of the invention.

FIG. **10** conceptually illustrates a 5×5 array camera that can be utilized to construct a depth map in accordance with an embodiment of the invention.

FIG. **11** is a flowchart illustrating a process for determining visibility based upon the photometric similarity of corresponding pixels in accordance with an embodiment of the invention.

FIG. **12** conceptually illustrates one of many virtual viewpoints that can be defined with respect to a 4×4 array camera in accordance with an embodiment of the invention.

FIG. **13** is a flowchart illustrating a process for generating a sparse depth map in accordance with an embodiment of the invention.

FIG. **14** conceptually illustrates a set of pixels that can be utilized as indicator pixels when generating a sparse depth map in accordance with an embodiment of the invention.

FIG. **15** is a flowchart illustrating a process for detecting textureless regions using the SNR surrounding a pixel in accordance with an embodiment of the invention.

FIG. **16** is a system for generating a depth map and visibility information in accordance with an embodiment of the invention.

FIG. **17** is a flowchart illustrating a process for synthesizing a higher resolution image from a plurality of lower resolution images captured from different viewpoints using super-resolution processing in accordance with an embodiment of the invention.

FIGS. **18**A and **18**B conceptually illustrate sources of noise in depth estimates.

FIGS. **18**C-**18**H conceptually illustrate the generation of a depth map and a confidence map from captured image data and the use of the confidence map to filter the depth map in accordance with an embodiment of the invention.

FIGS. **18**I-**18**N similarly conceptually illustrate the generation of a depth map and a confidence map from captured image data and the use of the confidence map to filter the depth map using close up images in accordance with an embodiment of the invention.

FIG. **19** is a plan view of a camera array with a plurality of imagers, according to one embodiment.

FIGS. **20**A through **20**E are plan views of camera arrays having different layouts of heterogeneous imagers, according to embodiments of the invention.

FIG. **21**A conceptually illustrates a 3×3 camera module patterned with a π filter group where red cameras are arranged horizontally and blue cameras are arranged vertically in accordance with an embodiment of the invention.

FIG. **21**B conceptually illustrates a 3×3 camera module patterned with a π filter group where red cameras are arranged vertically and blue cameras are arranged horizontally in accordance with an embodiment of the invention.

FIG. **22** conceptually illustrates a 4×4 camera module patterned with two π filter groups in accordance with an embodiment of the invention.

FIG. **23** conceptually illustrates a 4×4 camera module patterned with two π filter groups with two cameras that could each act as a reference camera in accordance with an embodiment of the invention.

FIG. **24** conceptually illustrates a 4×4 camera module patterned with π filter groups where nine cameras are utilized to capture image data used to synthesize frames of video in accordance with an embodiment of the invention.

FIG. **25** is a flow chart of a process for creating a light field image file including an image synthesized from light field image data and a depth map for the synthesized image generated using the light field image data in accordance with an embodiment of the invention.

FIG. **26** is a process for creating a light field image file that conforms to the JFIF standard and that includes an image encoded in accordance with the JPEG standard in accordance with an embodiment of the invention.

FIG. **27** illustrates an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **28** illustrates a "DZ Selection Descriptor" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **29** illustrates a "Depth Map, Camera Array and Auxiliary Maps Selection Descriptor" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **30** illustrates a "Depth Map, Camera Array and Auxiliary Maps Compression Descriptor" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **31** illustrates a "Depth Map Attributes" field within a "Depth Map Header" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **32** illustrates a "Depth Map Descriptor" field within a "Depth Map Header" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **33** illustrates a "Depth Map Data Descriptor" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **34** illustrates a "Camera Array Attributes" field within a "Camera Array Header" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **35** illustrates a "Camera Array Descriptor" field within a "Camera Array Header" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **36** illustrates an "Individual Camera Descriptor" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **37** illustrates "Individual Camera Data" within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **38** illustrates an "Individual Pixel Data Structure" within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **39** illustrates an "Auxiliary Map Descriptor" within an "Auxiliary Map Header" contained within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **40** illustrates an "Auxiliary Map Data Descriptor" within an APP9 Application marker segment of a light field image file that conforms to the JFIF standard in accordance with an embodiment of the invention.

FIG. **41** is a flow chart illustrating a process for rendering an image using a light field image file in accordance with an embodiment of the invention.

FIG. **42** is a flow chart illustrating a process for rendering an image using a light field image file that conforms to the JFIF standard and includes an image encoded in accordance with the JPEG standard and metadata describing the encoded image.

DETAILED DESCRIPTION

Turning now to the drawings, systems and methods for parallax detection and correction in images captured using array cameras are illustrated. Array cameras, such as those described in U.S. patent application Ser. No. 12/935,504 entitled "Capturing and Processing of Images using Monolithic Camera Array with Heterogeneous Imagers" to Venkataraman et al., can be utilized to capture light field images. In a number of embodiments, super-resolution processes such as those described in U.S. patent application Ser. No. 12/967,807 entitled "Systems and Methods for Synthesizing High Resolution Images Using Super-Resolution Processes" to Lelescu et al., are utilized to synthesize a higher resolution 2D image or a stereo pair of higher resolution 2D images from the lower resolution images in the light field captured by an array camera. The terms high or higher resolution and low or lower resolution are used here in a relative sense and not to indicate the specific resolutions of the images captured by the array camera. The disclosures of U.S. patent application Ser. No. 12/935,504 and U.S. patent application Ser. No. 12/967,807 are hereby incorporated by reference in their entirety.

Each two-dimensional (2D) image in a captured light field is from the viewpoint of one of the cameras in the array camera. Due to the different viewpoint of each of the cameras, parallax results in variations in the position of objects within the different images of the scene. Systems and methods in accordance with embodiments of the invention provide an accurate account of the pixel disparity as a result of parallax between the different cameras in the array, so that appropriate

scene-dependent geometric shifts can be applied to the pixels of the captured images when performing super-resolution processing.

A high resolution image synthesized using super-resolution processing is synthesized from a specific viewpoint that can be referred to as a reference viewpoint. The reference viewpoint can be from the viewpoint of one of the cameras in a camera array. Alternatively, the reference viewpoint can be an arbitrary virtual viewpoint where there is no physical camera. A benefit of synthesizing a high resolution image from the viewpoint of one of the cameras (as opposed to a virtual viewpoint) is that the disparity of the pixels in the light field can be determined with respect to the image in the light field captured from the reference viewpoint. When a virtual viewpoint is utilized, none of the captured image data is from the reference viewpoint and so the process instead relies solely on cameras away from the reference position to determine the best match.

Array cameras in accordance with many embodiments of the invention use the disparity between the pixels in the images within a light field to generate a depth map from the reference viewpoint. A depth map indicates the distance of scene objects from a reference viewpoint and can be utilized to determine scene dependent geometric corrections to apply to the pixels from each of the images within a captured light field to correct for disparity when performing super-resolution processing. In several embodiments, an initial depth map of the reference viewpoint is generated and as part of that process or as a subsequent process occluded pixels and/or other types of mismatched pixels are detected. The process of detecting pixels that are occluded can also be thought of as determining whether a pixel in an image captured from the reference viewpoint is visible in the image from a non-reference viewpoint. When a pixel in the image captured from the reference viewpoint is not visible in a second image, utilizing image data from the second image when determining the depth of the pixel in the reference image introduces error into the depth determination. Therefore, by detecting the pixels in the reference image that are occluded in one or more images in the light field, the accuracy of the depth map can be improved. In several embodiments, the initial depth map is updated by determining the depths of occluded pixels using image data captured from cameras in which the pixels are visible (i.e. not occluded). In a number of embodiments, the likely presence of occlusions and/or other sources of mismatched pixels can be detected during the process of generating an initial depth estimate and subsets of a set of images that correspond to different patterns of visibility within a scene can be used to determine a set of candidate depth estimates. The candidate depth of the subset of images having the most similar corresponding pixels can be used as the new depth estimate and the new depth estimate used to determine the visibility of the corresponding pixels in some or all of the remaining set of images.

A depth map from a reference viewpoint can be utilized to determine the scene dependent geometric shifts that are likely to have occurred in images captured from other viewpoints. These scene dependent geometric shifts can be utilized in super-resolution processing. In addition, the scene dependent geometric shifts can be utilized to refine the determinations of the visibility of pixels within the light field from the reference viewpoint. In a number of embodiments, the scene dependent geometric shifts are utilized to compare the similarity of pixels. Assuming the depth of a pixel from the reference viewpoint is correctly determined, then the similarity of the pixels is indicative of whether the pixel is visible. A similar pixel is likely to be the pixel observed from the reference viewpoint shifted due to disparity. If the pixels are dissimilar, then the pixel observed from the reference viewpoint is likely occluded in the second image. In many embodiments, visibility information is utilized in further updating depth maps. In several embodiments, visibility information is generated and provided along with the depth map for use in super-resolution processing.

In a number of embodiments, the computational complexity of generating depth maps is reduced by generating a sparse depth map that includes additional depth estimates in regions where additional depth information is desirable such as (but not limited to) regions involving depth transitions and/or regions containing pixels that are occluded in one or more images within the light field.

Many array cameras capture color information using different cameras (see for example the array cameras disclosed in U.S. patent application Ser. No. 12/935,504). In many embodiments, the viewpoint of a Green camera is utilized as the reference viewpoint. An initial depth map can be generated using the images captured by other Green cameras in the array camera and the depth map used to determine the visibility of Red, Green, and Blue pixels within the light field. In other embodiments, image data in multiple color channels can be utilized to perform depth estimation. In several embodiments, the similarity of corresponding pixels in each color channel is considered when estimating depth. In a number of embodiments, the similarity of sets of corresponding pixels in different color channels is also considered when estimating depth. Depth estimation using various cost functions that consider the similarity of corresponding pixels at specific depths in a single spectral channel, in multiple spectral channels, and/or across spectral channels in accordance with embodiments of the invention are discussed further below.

In several embodiments, the array camera can include one or more cameras that capture image data in multiple color channels. For example, an array camera may include one or more cameras that have a Bayer color filter pattern, in addition to or as an alternative to monochrome cameras. When the viewpoint of a camera that captures multiple color channels is used as the reference viewpoint for the purpose of generating a depth map, a depth map and visibility information can be determined for each color channel captured from the reference viewpoint. When a reference image contains information concerning multiple color channels, depth and visibility information can be more reliably created based upon the disparity of the pixels in the light field with respect to the reference image than by registering the pixels in one channel with respect to the depth and visibility of pixels in another color channel. A disadvantage of utilizing the viewpoint of a camera that captures image data in multiple color channels as a reference viewpoint is that the resolution of the depth information in each of the captured color channels is reduced relative to a camera that captures image data using the same number of pixels in a single channel. Accordingly, the configuration of the array camera and the selection of the viewpoint to utilize as the reference viewpoint typically depend upon the requirements of a specific application.

Once a depth map and visibility information are generated for the pixels in the light field, the depth map and visibility information can be provided to a super-resolution processing pipeline in accordance with embodiments of the invention to synthesize a higher resolution 2D image of the scene. The depth map can be utilized to correct for parallax between the different low resolution images and visibility information can be utilized during fusion to prevent the fusion of occluded pixels (i.e. pixels in an alternate view image that are not

visible from the reference viewpoint). In several embodiments, the process of generating a depth map also includes generating a confidence map that includes confidence metrics for the depth estimates in the depth map. In several embodiments, the depth metrics encode at least one confidence factor indicative of the reliability of the corresponding depth estimate. In a number of embodiments, the confidence metric includes at least a confidence factor based on the signal to noise ratio (SNR) in the region of the pixel location with which the depth estimate is associated, and a confidence factor based upon the number of pixels in a set of images that correspond to the pixel location with which the depth map is associated that were utilized to generate the depth estimate and/or are occluded. Systems and methods for detecting and correcting disparity in images captured by array cameras in accordance with embodiments of the invention are described below. Before discussing the detection and correction of parallax, however, various array cameras in accordance with embodiments of the invention are discussed.

Array Cameras

Embodiments relate to using a distributed approach to capturing images using a plurality of imagers of different imaging characteristics. Each imager may be spatially shifted from another imager in such a manner that an imager captures an image that us shifted by a sub-pixel amount with respect to another imager captured by another imager. Each imager may also include separate optics with different filters and operate with different operating parameters (e.g., exposure time). Distinct images generated by the imagers are processed to obtain an enhanced image. Each imager may be associated with an optical element fabricated using wafer level optics (WLO) technology.

A sensor element or pixel refers to an individual light sensing element in a camera array. The sensor element or pixel includes, among others, traditional CIS (CMOS Image Sensor), CCD (charge-coupled device), high dynamic range pixel, multispectral pixel and various alternatives thereof.

An imager refers to a two dimensional array of pixels. The sensor elements of each imager have similar physical properties and receive light through the same optical component. Further, the sensor elements in the each imager may be associated with the same color filter.

A camera array refers to a collection of imagers designed to function as a unitary component. The camera array may be fabricated on a single chip for mounting or installing in various devices.

An array of camera arrays refers to an aggregation of two or more camera arrays. Two or more camera arrays may operate in conjunction to provide extended functionality over a single camera array.

Image characteristics of an imager refer to any characteristics or parameters of the imager associated with capturing of images. The imaging characteristics may include, among others, the size of the imager, the type of pixels included in the imager, the shape of the imager, filters associated with the imager, the exposure time of the imager, aperture size associated with the imager, the configuration of the optical element associated with the imager, gain of the imager, the resolution of the imager, and operational timing of the imager.

Structure of Camera array

FIG. **19** is a plan view of a camera array **1900** with imagers **1**A through NM, according to one embodiment. The camera array **1900** is fabricated on a semiconductor chip to include a plurality of imagers **1**A through NM. Each of the imagers **1**A through NM may include a plurality of pixels (e.g., 0.32 Mega pixels). In one embodiment, the imagers **1**A through NM are arranged into a grid format as illustrated in FIG. **19**. In other embodiments, the imagers are arranged in a non-grid format. For example, the imagers may be arranged in a circular pattern, zigzagged pattern or scattered pattern.

The camera array may include two or more types of heterogeneous imagers, each imager including two or more sensor elements or pixels. Each one of the imagers may have different imaging characteristics. Alternatively, there may be two or more different types of imagers where the same type of imagers shares the same imaging characteristics.

In one embodiment, each imager **1**A through NM has its own filter and/or optical element (e.g., lens). Specifically, each of the imagers **1**A through NM or a group of imagers may be associated with spectral color filters to receive certain wavelengths of light. Example filters include a traditional filter used in the Bayer pattern (R, G, B or their complements C, M, Y), an IR-cut filter, a near-IR filter, a polarizing filter, and a custom filter to suit the needs of hyper-spectral imaging. Some imagers may have no filter to allow reception of both the entire visible spectra and near-IR, which increases the imager's signal-to-noise ratio. The number of distinct filters may be as large as the number of imagers in the camera array. Further, each of the imagers **1**A through NM or a group of imagers may receive light through lens having different optical characteristics (e.g., focal lengths) or apertures of different sizes.

In one embodiment, the camera array includes other related circuitry. The other circuitry may include, among others, circuitry to control imaging parameters and sensors to sense physical parameters. The control circuitry may control imaging parameters such as exposure times, gain, and black level offset. The sensor may include dark pixels to estimate dark current at the operating temperature. The dark current may be measured for on-the-fly compensation for any thermal creep that the substrate may suffer from.

In one embodiment, the circuit for controlling imaging parameters may trigger each imager independently or in a synchronized manner. The start of the exposure periods for the various imagers in the camera array (analogous to opening a shutter) may be staggered in an overlapping manner so that the scenes are sampled sequentially while having several imagers being exposed to light at the same time. In a conventional video camera sampling a scene at N exposures per second, the exposure time per sample is limited to 1/N seconds. With a plurality of imagers, there is no such limit to the exposure time per sample because multiple imagers may be operated to capture images in a staggered manner.

Each imager can be operated independently. Entire or most operations associated with each individual imager may be individualized. In one embodiment, a master setting is programmed and deviation (i.e., offset or gain) from such master setting is configured for each imager. The deviations may reflect functions such as high dynamic range, gain settings, integration time settings, digital processing settings or combinations thereof. These deviations can be specified at a low level (e.g., deviation in the gain) or at a higher level (e.g., difference in the ISO number, which is then automatically translated to deltas for gain, integration time, or otherwise as specified by context/master control registers) for the particular camera array. By setting the master values and deviations from the master values, higher levels of control abstraction can be achieved to facilitate simpler programming model for many operations. In one embodiment, the parameters for the imagers are arbitrarily fixed for a target application. In another embodiment, the parameters are configured to allow a high degree of flexibility and programmability.

In one embodiment, the camera array is designed as a drop-in replacement for existing camera image sensors used

in cell phones and other mobile devices. For this purpose, the camera array may be designed to be physically compatible with conventional image sensors of approximately the same resolution although the achieved resolution of the camera array may exceed conventional image sensors in many photographic situations. Taking advantage of the increased performance, the camera array of the embodiment may include fewer pixels to obtain equal or better quality images compared to conventional image sensors. Alternatively, the size of the pixels in the imager may be reduced compared to pixels in conventional image sensors while achieving comparable results.

In order to match the raw pixel count of a conventional image sensor without increasing silicon area, the logic overhead for the individual imagers is preferably constrained in the silicon area. In one embodiment, much of the pixel control logic is a single collection of functions common to all or most of the imagers with a smaller set of functions applicable each imager. In this embodiment, the conventional external interface for the imager may be used because the data output does not increase significantly for the imagers.

The number of imagers in the camera array may be determined based on, among other factors, (i) resolution, (ii) parallax, (iii) sensitivity, and (iv) dynamic range. A first factor for the size of imager is the resolution. From a resolution point of view, the preferred number of the imagers ranges from 2×2 to 6×6 because an array size of larger than 6×6 is likely to destroy frequency information that cannot be recreated by the super-resolution process. For example, 8 Megapixel resolution with 2×2 imager will require each imager to have 2 Megapixels. Similarly, 8 Megapixel resolution with a 5×5 array will require each imager to have 0.32 Megapixels.

A second factor that may constrain the number of imagers is the issue of parallax and occlusion. With respect to an object captured in an image, the portion of the background scene that is occluded from the view of the imager is called as "occlusion set." When two imagers capture the object from two different locations, the occlusion set of each imager is different. Hence, there may be scene pixels captured by one imager but not the other. To resolve this issue of occlusion, it is desirable to include a certain minimal set of imagers for a given type of imager.

A third factor that may put a lower bound on the number of imagers is the issue of sensitivity in low light conditions. To improve low light sensitivity, imagers for detecting near-IR spectrum may be needed. The number of imagers in the camera array may need to be increased to accommodate such near-IR imagers.

A fourth factor in determining the size of the imager is dynamic range. To provide dynamic range in the camera array, it is advantageous to provide several imagers of the same filter type (chroma or luma). Each imager of the same filter type may then be operated with different exposures simultaneously. The images captured with different exposures may be processed to generate a high dynamic range image.

Based on these factors, the preferred number of imagers is 2×2 to 6×6. 4×4 and 5×5 configurations are more preferable than 2×2 and 3×3 configurations because the former are likely to provide sufficient number of imagers to resolve occlusion issues, increase sensitivity and increase the dynamic range. At the same time, the computational load required to recover resolution from these array sizes will be modest in comparison to that required in the 6×6 array. Arrays larger than 6×6 may, however, be used to provide additional features such as optical zooming and multispectral imaging.

Another consideration is the number of imagers dedicated to luma sampling. By ensuring that the imagers in the array dedicated to near-IR sampling do not reduce the achieved resolution, the information from the near-IR images is added to the resolution captured by the luma imagers. For this purpose, at least 50% of the imagers may be used for sampling the luma and/or near-IR spectra. In one embodiment with 4×4 imagers, 4 imagers samples luma, 4 imagers samples near-IR, and the remaining 8 imagers samples two chroma (Red and Blue). In another embodiment with 5×5 imagers, 9 imagers samples luma, 8 imagers samples near-IR, and the remaining 8 imagers samples two chroma (Red and Blue). Further, the imagers with these filters may be arranged symmetrically within the camera array to address occlusion due to parallax.

In one embodiment, the imagers in the camera array are spatially separated from each other by a predetermined distance. By increasing the spatial separation, the parallax between the images captured by the imagers may be increased. The increased parallax is advantageous where more accurate distance information is important. Separation between two imagers may also be increased to approximate the separation of a pair of human eyes. By approximating the separation of human eyes, a realistic stereoscopic 3D image may be provided to present the resulting image on an appropriate 3D display device.

In one embodiment, multiple camera arrays are provided at different locations on a device to overcome space constraints. One camera array may be designed to fit within a restricted space while another camera array may be placed in another restricted space of the device. For example, if a total of 20 imagers are required but the available space allows only a camera array of 1×10 imagers to be provided on either side of a device, two camera arrays each including 10 imagers may be placed on available space at both sides of the device. Each camera array may be fabricated on a substrate and be secured to a motherboard or other parts of a device. The images collected from multiple camera arrays may be processed to generate images of desired resolution and performance.

A design for a single imager may be applied to different camera arrays each including other types of imagers. Other variables in the camera array such as spatial distances, color filters and combination with the same or other sensors may be modified to produce a camera array with differing imaging characteristics. In this way, a diverse mix of camera arrays may be produced while maintaining the benefits from economies of scale.

FIGS. **20**A through **20**E illustrate various configurations of imagers for obtaining a high resolution image through a super-resolution process, according to embodiments of the present invention. In FIGS. **20**A through **20**E, "R" represents an imager having a red filter, "G" represents a imager having a green filter, "B" represents an imager having a blue filter, "P" represents a polychromatic imager having sensitivity across the entire visible spectra and near-IR spectrum, and "I" represents an imager having a near-IR filter. The polychromatic imager may sample image from all parts of the visible spectra and the near-IR region (i.e., from 650 nm to 800 nm). In the embodiment of FIG. **20**A, the center columns and rows of the imagers include polychromatic imagers. The remaining areas of the camera array are filled with imagers having green filters, blue filters, and red filters. The embodiment of FIG. **20**A does not include any imagers for detecting near-IR spectrum alone.

The embodiment of FIG. **20**B has a configuration similar to conventional Bayer filter mapping. This embodiment does not include any polychromatic imagers or near-IR imagers. As described above in detail with reference to FIG. **19**, the

embodiment of FIG. **20**B is different from conventional Bayer filter configuration in that each color filter is mapped to each imager instead of being mapped to an individual pixel.

FIG. **20**C illustrates an embodiment where the polychromatic imagers form a symmetric checkerboard pattern. FIG. **20**D illustrates an embodiment where four near-IR imagers are provided. FIG. **20**E illustrates an embodiment with irregular mapping of imagers. The embodiments of FIGS. **20**A through **20**E are merely illustrative and various other layouts of imagers can also be used.

The use of polychromatic imagers and near-IR imagers is advantageous because these sensors may capture high quality images in low lighting conditions. The images captured by the polychromatic imager or the near-IR imager are used to denoise the images obtained from regular color imagers.

The premise of increasing resolution by aggregating multiple low resolution images is based on the fact that the different low resolution images represent slightly different viewpoints of the same scene. If the LR images are all shifted by integer units of a pixel, then each image contains essentially the same information. Therefore, there is no new information in LR images that can be used to create the HR image. In the imagers according to embodiments, the layout of the imagers may be preset and controlled so that each imager in a row or a column is a fixed sub-pixel distance from its neighboring imagers. The wafer level manufacturing and packaging process allows accurate formation of imagers to attain the sub-pixel precisions required for the super-resolution processing.

An issue of separating the spectral sensing elements into different imagers is parallax caused by the physical separation of the imagers. By ensuring that the imagers are symmetrically placed, at least two imagers can capture the pixels around the edge of a foreground object. In this way, the pixels around the edge of a foreground object may be aggregated to increase resolution as well as avoiding any occlusions. Another issue related to parallax is the sampling of color. The issue of sampling the color may be reduced by using parallax information in the polychromatic imagers to improve the accuracy of the sampling of color from the color filtered imagers.

In one embodiment, near-IR imagers are used to determine relative luminance differences compared to a visible spectra imager. Objects have differing material reflectivity results in differences in the images captured by the visible spectra and the near-IR spectra. At low lighting conditions, the near-IR imager exhibits a higher signal to noise ratios. Therefore, the signals from the near-IR sensor may be used to enhance the luminance image. The transferring of details from the near-IR image to the luminance image may be performed before aggregating spectral images from different imagers through the super-resolution process. In this way, edge information about the scene may be improved to construct edge-preserving images that can be used effectively in the super-resolution process.

Image Fusion of Color Images with Near-IR Images

The spectral response of CMOS imagers is typically very good in the near-IR regions covering 650 nm to 800 nm and reasonably good between 800 nm and 1000 nm. Although near-IR images having no chroma information, information in this spectral region is useful in low lighting conditions because the near-IR images are relatively free of noise. Hence, the near-IR images may be used to denoise color images under the low lighting conditions.

In one embodiment, an image from a near-IR imager is fused with another image from a visible light imager. Before proceeding with the fusion, a registration is performed between the near-IR image and the visible light image to resolve differences in viewpoints. The registration process may be performed in an offline, one-time, processing step. After the registration is performed, the luminance information on the near-IR image is interpolated to grid points that correspond to each grid point on the visible light image.

After the pixel correspondence between the near-IR image and the visible light image is established, denoising and detail transfer process may be performed. The denoising process allows transfer of signal information from the near-IR image to the visible light image to improve the overall SNR of the fusion image. The detail transfer ensures that edges in the near-IR image and the visible light image are preserved and accentuated to improve the overall visibility of objects in the fused image.

In one embodiment, a near-IR flash may serve as a near-IR light source during capturing of an image by the near-IR imagers. Using the near-IR flash is advantageous, among other reasons, because (i) the harsh lighting on objects of interest may be prevented, (ii) ambient color of the object may be preserved, and (iii) red-eye effect may be prevented.

In one embodiment, a visible light filter that allows only near-IR rays to pass through is used to further optimize the optics for near-IR imaging. The visible light filter improves the near-IR optics transfer function because the light filter results in sharper details in the near-IR image. The details may then be transferred to the visible light images using a dual bilateral filter.

Dynamic Range Determination by Differing Exposures at Imagers

An auto-exposure (AE) algorithm is important to obtaining an appropriate exposure for the scene to be captured. The design of the AE algorithm affects the dynamic range of captured images. The AE algorithm determines an exposure value that allows the acquired image to fall in the linear region of the camera array's sensitivity range. The linear region is preferred because a good signal-to-noise ratio is obtained in this region. If the exposure is too low, the picture becomes under-saturated while if the exposure is too high the picture becomes over-saturated. In conventional cameras, an iterative process is taken to reduce the difference between measured picture brightness and previously defined brightness below a threshold. This iterative process requires a large amount of time for convergence, and sometimes results in an unacceptable shutter delay.

In one embodiment, the picture brightness of images captured by a plurality of imagers is independently measured. Specifically, a plurality of imagers are set to capturing images with different exposures to reduce the time for computing the adequate exposure. For example, in a camera array with 5×5 imagers where 8 luma imagers and 9 near-IR imagers are provided, each of the imagers may be set with different exposures. The near-IR imagers are used to capture low-light aspects of the scene and the luma imagers are used to capture the high illumination aspects of the scene. This results in a total of 17 possible exposures. If exposure for each imager is offset from an adjacent imager by a factor of 2, for example, a maximum dynamic range of 2^{17} or 102 dB can be captured. This maximum dynamic range is considerably higher than the typical 48 dB attainable in a conventional camera with 8 bit image outputs.

At each time instant, the responses (under-exposed, over-exposed or optimal) from each of the multiple imagers are analyzed based on how many exposures are needed at the subsequent time instant. The ability to query multiple exposures simultaneously in the range of possible exposures accelerates the search compared to the case where only one

exposure is tested at once. By reducing the processing time for determining the adequate exposure, shutter delays and shot-to-shot lags may be reduced.

In one embodiment, the HDR image is synthesized from multiple exposures by combining the images after linearizing the imager response for each exposure. The images from the imagers may be registered before combining to account for the difference in the viewpoints of the imagers.

In one embodiment, at least one imager includes HDR pixels to generate HDR images. HDR pixels are specialized pixels that capture high dynamic range scenes. Although HDR pixels show superior performances compared to other pixels, HDR pixels show poor performance at low lighting conditions in comparison with near-IR imagers. To improve performance at low lighting conditions, signals from the near-IR imagers may be used in conjunction with the signal from the HDR imager to attain better quality images across different lighting conditions.

In one embodiment, an HDR image is obtained by processing images captured by multiple imagers by processing. The ability to capture multiple exposures simultaneously using the imager is advantageous because artifacts caused by motion of objects in the scene can be mitigated or eliminated.

Hyperspectral Imaging by Multiple Imagers

In one embodiment, a multi-spectral image is rendered by multiple imagers to facilitate the segmentation or recognition of objects in a scene. Because the spectral reflectance coefficients vary smoothly in most real world objects, the spectral reflectance coefficients may be estimated by capturing the scene in multiple spectral dimensions using imagers with different color filters and analyzing the captured images using Principal Components Analysis (PCA).

In one embodiment, half of the imagers in the camera array are devoted to sampling in the basic spectral dimensions (R, G, and B) and the other half of the imagers are devoted to sampling in a shifted basic spectral dimensions (R', G', and B'). The shifted basic spectral dimensions are shifted from the basic spectral dimensions by a certain wavelength (e.g., 10 nm).

In one embodiment, pixel correspondence and non-linear interpolation is performed to account for the sub-pixel shifted views of the scene. Then the spectral reflectance coefficients of the scene are synthesized using a set of orthogonal spectral basis functions. The basis functions are eigenvectors derived by PCA of a correlation matrix and the correlation matrix is derived from a database storing spectral reflectance coefficients measured by, for example, Munsell color chips (a total of 1257) representing the spectral distribution of a wide range of real world materials to reconstruct the spectrum at each point in the scene.

At first glance, capturing different spectral images of the scene through different imagers in the camera array appears to trade resolution for higher dimensional spectral sampling. However, some of the lost resolution may be recovered. The multiple imagers sample the scene over different spectral dimensions where each sampling grid of each imager is offset by a sub-pixel shift from the others. In one embodiment, no two sampling grid of the imager overlap. That is, the superposition of all the sampling grids from all the imagers forms a dense, possibly non-uniform, montage of points. Scattered data interpolation methods may be used to determine the spectral density at each sample point in this non-uniform montage for each spectral image. In this way, a certain amount of resolution lost in the process of sampling the scene using different spectral filters may be recovered.

As described above, image segmentation and object recognition are facilitated by determining the spectral reflectance coefficients of the object. The situation often arises in security applications wherein a network of cameras is used to track an object as it moves from the operational zone of one camera to another. Each zone may have its own unique lighting conditions (fluorescent, incandescent, D65, etc.) that may cause the object to have a different appearance in each image captured by different cameras. If these cameras capture the images in a hyper-spectral mode, all images may be converted to the same illuminant to enhance object recognition performance.

In one embodiment, camera arrays with multiple imagers are used for providing medical diagnostic images. Full spectral digitized images of diagnostic samples contribute to accurate diagnosis because doctors and medical personnel can place higher confidence in the resulting diagnosis. The imagers in the camera arrays may be provided with color filters to provide full spectral data. Such camera array may be installed on cell phones to capture and transmit diagnostic information to remote locations. Further, the camera arrays including multiple imagers may provide images with a large depth of field to enhance the reliability of image capture of wounds, rashes, and other symptoms.

In one embodiment, a small imager (including, for example, 20-500 pixels) with a narrow spectral bandpass filters is used to produce a signature of the ambient and local light sources in a scene. By using the small imager, the exposure and white balance characteristics may be determined more accurately at a faster speed. The spectral bandpass filters may be ordinary color filters or diffractive elements of a bandpass width adequate to allow the number of camera arrays to cover the visible spectrum of about 400 nm. These imagers may run at a much higher frame rate and obtain data (which may or may not be used for its pictorial content) for processing into information to control the exposure and white balance of other larger imagers in the same camera array. The small imagers may also be interspersed within the camera array.

Optical Zoom Implemented Using Multiple Imagers

In one embodiment, a subset of imagers in the camera array includes telephoto lenses. The subset of imagers may have other imaging characteristics same as imagers with non-telephoto lenses. Images from this subset of imagers are combined and super-resolution processed to form a super-resolution telephoto image. In another embodiment, the camera array includes two or more subsets of imagers equipped with lenses of more than two magnifications to provide differing zoom magnifications.

Embodiments of the camera arrays may achieve its final resolution by aggregating images through super-resolution. Taking an example of providing 5×5 imagers with a 3× optical zoom feature, if 17 imagers are used to sample the luma (G) and 8 imagers are used to sample the chroma (R and B), 17 luma imagers allow a resolution that is four times higher than what is achieved by any single imager in the set of 17 imagers. If the number of the imager is increased from 5×5 to 6×6, an addition of 11 extra imagers becomes available. In comparison with the 8 Megapixel conventional image sensor fitted with a 3× zoom lens, a resolution that is 60% of the conventional image sensor is achieved when 8 of the additional 11 imagers are dedicated to sampling luma (G) and the remaining 3 imagers are dedicated to chroma (R and B) and near-IR sampling at 3× zoom. This considerably reduces the chroma sampling (or near-IR sampling) to luma sampling ratio. The reduced chroma to luma sampling ratio is somewhat offset by using the super-resolved luma image at 3× zoom as a recognition prior on the chroma (and near-IR) image to resample the chroma image at a higher resolution.

With 6×6 imagers, a resolution equivalent to the resolution of conventional image sensor is achieved at 1× zoom. At 3× zoom, a resolution equivalent to about 60% of conventional image sensor outfitted with a 3× zoom lens is obtained by the same imagers. Also, there is a decrease in luma resolution at 3× zoom compared with conventional image sensors with resolution at 3× zoom. The decreased luma resolution, however, is offset by the fact that the optics of conventional image sensor has reduced efficiency at 3× zoom due to crosstalk and optical aberrations.

The zoom operation achieved by multiple imagers has the following advantages. First, the quality of the achieved zoom is considerably higher than what is achieved in the conventional image sensor due to the fact that the lens elements may be tailored for each change in focal length. In conventional image sensors, optical aberrations and field curvature must be corrected across the whole operating range of the lens, which is considerably harder in a zoom lens with moving elements than in a fixed lens element where only aberrations for a fixed focal length need to be corrected. Additionally, the fixed lens in the imagers has a fixed chief ray angle for a given height, which is not the case with conventional image sensor with a moving zoom lens. Second, the imagers allow simulation of zoom lenses without significantly increasing the optical track height. The reduced height allows implementation of thin modules even for camera arrays with zooming capability.

The overhead required to support a certain level of optical zoom in camera arrays according to some embodiments is tabulated in Table 2.

TABLE 2

No. of Imagers in Camera	No. of Luma Imagers at different Zoom levels			No. of Chroma Imagers at different Zoom Levels		
array	1×	2×	3×	1×	2×	3×
25	17	0	0	8	0	0
36	16	0	8	8	0	4

In one embodiment, the pixels in the images are mapped onto an output image with a size and resolution corresponding to the amount of zoom desired in order to provide a smooth zoom capability from the widest-angle view to the greatest-magnification view. Assuming that the higher magnification lenses have the same center of view as the lower magnification lenses, the image information available is such that a center area of the image has a higher resolution available than the outer area. In the case of three or more distinct magnifications, nested regions of different resolution may be provided with resolution increasing toward the center.

An image with the most telephoto effect has a resolution determined by the super-resolution ability of the imagers equipped with the telephoto lenses. An image with the widest field of view can be formatted in at least one of two following ways. First, the wide field image may be formatted as an image with a uniform resolution where the resolution is determined by the super-resolution capability of the set of imagers having the wider-angle lenses. Second, the wide field image is formatted as a higher resolution image where the resolution of the central part of the image is determined by the super-resolution capability of the set of imagers equipped with telephoto lenses. In the lower resolution regions, information from the reduced number of pixels per image area is interpolated smoothly across the larger number of "digital" pixels. In such an image, the pixel information may be processed and interpolated so that the transition from higher to lower resolution regions occurs smoothly.

In one embodiment, zooming is achieved by inducing a barrel-like distortion into some, or all, of the array lens so that a disproportionate number of the pixels are dedicated to the central part of each image. In this embodiment, every image has to be processed to remove the barrel distortion. To generate a wide angle image, pixels closer to the center are sub-sampled relative to outer pixels are super-sampled. As zooming is performed, the pixels at the periphery of the imagers are progressively discarded and the sampling of the pixels nearer the center of the imager is increased.

In one embodiment, mipmap filters are built to allow images to be rendered at a zoom scale that is between the specific zoom range of the optical elements (e.g., 1× and 3× zoom scales of the camera array). Mipmaps are a precalculated optimized set of images that accompany a baseline image. A set of images associated with the 3× zoom luma image can be created from a baseline scale at 3× down to 1×. Each image in this set is a version of the baseline 3× zoom image but at a reduced level of detail. Rendering an image at a desired zoom level is achieved using the mipmap by (i) taking the image at 1× zoom, and computing the coverage of the scene for the desired zoom level (i.e., what pixels in the baseline image needs to be rendered at the requested scale to produce the output image), (ii) for each pixel in the coverage set, determine if the pixel is in the image covered by the 3× zoom luma image, (iii) if the pixel is available in the 3× zoom luma image, then choose the two closest mipmap images and interpolate (using smoothing filter) the corresponding pixels from the two mipmap images to produce the output image, and (iv) if the pixel is unavailable in the 3× zoom luma image, then choose the pixel from the baseline 1× luma image and scale up to the desired scale to produce the output pixel. By using mipmaps, smooth optical zoom may be simulated at any point between two given discrete levels (i.e., 1× zoom and 3× zoom).

Capturing Video Images

In one embodiment, the camera array generates high frame image sequences. The imagers in the camera array can operate independently to capture images. Compared to conventional image sensors, the camera array may capture images at the frame rate up to N time (where N is the number of imagers). Further, the frame period for each imager may overlap to improve operations under low-light conditions. To increase the resolution, a subset of imagers may operate in a synchronized manner to produce images of higher resolution. In this case, the maximum frame rate is reduced by the number of imagers operating in a synchronized manner. The high-speed video frame rates can enables slow-motion video playback at a normal video rate.

In one example, two luma imagers (green imagers or near-IR imagers), two blue imagers and two green imagers are used to obtain high-definition 1080p images. Using permutations of four luma imagers (two green imagers and two near-IR imagers or three green imagers and one near-IR imager) together with one blue imager and one red imager, the chroma imagers can be upsampled to achieve 120 frames/sec for 1080p video. For higher frame rate imaging devices, the number of frame rates can be scaled up linearly. For Standard-Definition (480p) operation, a frame rate of 240 frames/sec may be achieved using the same camera array.

Conventional imaging devices with a high-resolution image sensor (e.g., 8 Megapixels) use binning or skipping to capture lower resolution images (e.g., 1080p30, 720p30 and 480p30). In binning, rows and columns in the captured images are interpolated in the charge, voltage or pixel

domains in order to achieve the target video resolutions while reducing the noise. In skipping, rows and columns are skipped in order to reduce the power consumption of the sensor. Both of these techniques result in reduced image quality.

In one embodiment, the imagers in the camera arrays are selectively activated to capture a video image. For example, 9 imagers (including one near-IR imager) may be used to obtain 1080p (1920×1080 pixels) images while 6 imagers (including one near-IR imager) may be used to obtain 720p (1280×720 pixels) images or 4 imagers (including one near-IR imager) may be used to obtain 480p (720×480 pixels) images. Because there is an accurate one-to-one pixel correspondence between the imager and the target video images, the resolution achieved is higher than traditional approaches. Further, since only a subset of the imagers is activated to capture the images, significant power savings can also be achieved. For example, 60% reduction in power consumption is achieved in 1080p and 80% of power consumption is achieved in 480p.

Using the near-IR imager to capture video images is advantageous because the information from the near-IR imager may be used to denoise each video image. In this way, the camera arrays of embodiments exhibit excellent low-light sensitivity and can operate in extremely low-light conditions. In one embodiment, super-resolution processing is performed on images from multiple imagers to obtain higher resolution video imagers. The noise-reduction characteristics of the super-resolution process along with fusion of images from the near-IR imager results in a very low-noise images.

In one embodiment, high-dynamic-range (HDR) video capture is enabled by activating more imagers. For example, in a 5×5 camera array operating in 1080p video capture mode, there are only 9 cameras active. A subset of the 16 cameras may be overexposed and underexposed by a stop in sets of two or four to achieve a video output with a very high dynamic range.

Other Applications for Multiple Imagers

In one embodiment, the multiple imagers are used for estimating distance to an object in a scene. Since information regarding the distance to each point in an image is available in the camera array along with the extent in x and y coordinates of an image element, the size of an image element may be determined. Further, the absolute size and shape of physical items may be measured without other reference information. For example, a picture of a foot can be taken and the resulting information may be used to accurately estimate the size of an appropriate shoe.

In one embodiment, reduction in depth of field is simulated in images captured by the camera array using distance information. The camera arrays according to the present invention produce images with greatly increased depth of field. The long depth of field, however, may not be desirable in some applications. In such case, a particular distance or several distances may be selected as the "in best focus" distance(s) for the image and based on the distance (z) information from parallax information, the image can be blurred pixel-by-pixel using, for example, a simple Gaussian blur. In one embodiment, the depth map obtained from the camera array is utilized to enable a tone mapping algorithm to perform the mapping using the depth information to guide the level, thereby emphasizing or exaggerating the 3D effect.

In one embodiment, apertures of different sizes are provided to obtain aperture diversity. The aperture size has a direct relationship with the depth of field. In miniature cameras, however, the aperture is generally made as large as possible to allow as much light to reach the camera array. Different imagers may receive light through apertures of different sizes. For imagers to produce a large depth of field, the aperture may be reduced whereas other imagers may have large apertures to maximize the light received. By fusing the images from sensor images of different aperture sizes, images with large depth of field may be obtained without sacrificing the quality of the image.

In one embodiment, the camera array according to the present invention refocuses based on images captured from offsets in viewpoints. Unlike a conventional plenoptic camera, the images obtained from the camera array of the present invention do not suffer from the extreme loss of resolution. The camera array according to the present invention, however, produces sparse data points for refocusing compared to the plenoptic camera. In order to overcome the sparse data points, interpolation may be performed to refocus data from the spare data points.

In one embodiment, each imager in the camera array has a different centroid. That is, the optics of each imager are designed and arranged so that the fields of view for each imager slightly overlap but for the most part constitute distinct tiles of a larger field of view. The images from each of the tiles are panoramically stitched together to render a single high-resolution image.

In one embodiment, camera arrays may be formed on separate substrates and mounted on the same motherboard with spatial separation. The lens elements on each imager may be arranged so that the corner of the field of view slightly encompasses a line perpendicular to the substrate. Thus, if four imagers are mounted on the motherboard with each imager rotated 90 degrees with respect to another imager, the fields of view will be four slightly overlapping tiles. This allows a single design of WLO lens array and imager chip to be used to capture different tiles of a panoramic image.

In one embodiment, one or more sets of imagers are arranged to capture images that are stitched to produce panoramic images with overlapping fields of view while another imager or sets of imagers have a field of view that encompasses the tiled image generated. This embodiment provides different effective resolution for imagers with different characteristics. For example, it may be desirable to have more luminance resolution than chrominance resolution. Hence, several sets of imagers may detect luminance with their fields of view panoramically stitched. Fewer imagers may be used to detect chrominance with the field of view encompassing the stitched field of view of the luminance imagers.

In one embodiment, the camera array with multiple imagers is mounted on a flexible motherboard such that the motherboard can be manually bent to change the aspect ratio of the image. For example, a set of imagers can be mounted in a horizontal line on a flexible motherboard so that in the quiescent state of the motherboard, the fields of view of all of the imagers are approximately the same. If there are four imagers, an image with double the resolution of each individual imager is obtained so that details in the subject image that are half the dimension of details that can be resolved by an individual imager. If the motherboard is bent so that it forms part of a vertical cylinder, the imagers point outward. With a partial bend, the width of the subject image is doubled while the detail that can be resolved is reduced because each point in the subject image is in the field of view of two rather than four imagers. At the maximum bend, the subject image is four times wider while the detail that can be resolved in the subject is further reduced.

Array Camera Architecture

Array cameras in accordance with embodiments of the invention can include a camera module including an array of cameras and a processor configured to read out and process

image data from the camera module to synthesize images. An array camera in accordance with an embodiment of the invention is illustrated in FIG. **1**. The array camera **100** includes a camera module **102** with an array of individual cameras **104** where an array of individual cameras refers to a plurality of cameras in a particular arrangement, such as (but not limited to) the square arrangement utilized in the illustrated embodiment. The camera module **102** is connected to the processor **108**. The processor is also configured to communicate with one or more different types of memory **110** that can be utilized to store image data and/or contain machine readable instructions utilized to configure the processor to perform processes including (but not limited to) the various processes described below. In many embodiments, the memory contains an image processing application that is configured to process a light field comprising a plurality of images to generate a depth map(s), a visibility map(s), a confidence map(s), and/or a higher resolution image(s) using any of the processes outlined in detail below. As is discussed further below, a depth map typically provides depth estimates for pixels in an image from a reference viewpoint (e.g. a higher resolution image synthesized from a reference viewpoint). A variety of visibility maps can be generated as appropriate to the requirements of specific applications including (but not limited to) visibility maps indicating whether pixel locations in a reference image are visible in specific images within a light field, visibility maps indicating whether specific pixels in an image within the light field are visible from the reference viewpoint, and visibility maps indicating whether a pixel visible in one alternate view image is visible in another alternate view image. In other embodiments, any of a variety of applications can be stored in memory and utilized to process image data using the processes described herein. In several embodiments, processes in accordance with embodiments of the invention can be implemented in hardware using an application specific integration circuit, and/or a field programmable gate array, or implemented partially in hardware and software.

Processors **108** in accordance with many embodiments of the invention are configured using appropriate software to take the image data within the light field and synthesize one or more high resolution images. In several embodiments, the high resolution image is synthesized from a reference viewpoint, typically that of a reference focal plane **104** within the sensor **102**. In many embodiments, the processor is able to synthesize an image from a virtual viewpoint, which does not correspond to the viewpoints of any of the focal planes **104** in the sensor **102**. The images in the light field will include a scene-dependent disparity due to the different fields of view of the focal planes used to capture the images. Processes for detecting and correcting for disparity are discussed further below. Although a specific array camera architecture is illustrated in FIG. **1**, alternative architectures can also be utilized in accordance with embodiments of the invention.

Array Camera Modules

Array camera modules in accordance with embodiments of the invention can be constructed from an imager array or sensor including an array of focal planes and an optic array including a lens stack for each focal plane in the imager array. Sensors including multiple focal planes are discussed in U.S. patent application Ser. No. 13/106,797 entitled "Architectures for System on Chip Array Cameras", to Pain et al., the disclosure of which is incorporated herein by reference in its entirety. Light filters can be used within each optical channel formed by the lens stacks in the optic array to enable different cameras within an array camera module to capture image data with respect to different portions of the electromagnetic spectrum (i.e. within different spectral channels).

An array camera module in accordance with an embodiment of the invention is illustrated in FIG. **1**A. The array camera module **150** includes an imager array **152** including an array of focal planes **154** along with a corresponding optic array **156** including an array of lens stacks **158**. Within the array of lens stacks, each lens stack **158** creates an optical channel that forms an image of the scene on an array of light sensitive pixels within a corresponding focal plane **154**. Each pairing of a lens stack **158** and focal plane **154** forms a single camera **104** within the camera module. Each pixel within a focal plane **154** of a camera **104** generates image data that can be sent from the camera **104** to the processor **108**. In many embodiments, the lens stack within each optical channel is configured so that pixels of each focal plane **158** sample the same object space or region within the scene. In several embodiments, the lens stacks are configured so that the pixels that sample the same object space do so with sub-pixel offsets to provide sampling diversity that can be utilized to recover increased resolution through the use of super-resolution processes. The term sampling diversity refers to the fact that the images from different viewpoints sample the same object in the scene but with slight sub-pixel offsets. By processing the images with sub-pixel precision, additional information encoded due to the sub-pixel offsets can be recovered when compared to simply sampling the object space with a single image.

In the illustrated embodiment, the focal planes are configured in a 5×5 array. Each focal plane **154** on the sensor is capable of capturing an image of the scene. Typically, each focal plane includes a plurality of rows of pixels that also forms a plurality of columns of pixels, and each focal plane is contained within a region of the imager that does not contain pixels from another focal plane. In many embodiments, image data capture and readout of each focal plane can be independently controlled. In this way, image capture settings including (but not limited to) the exposure times and analog gains of pixels within a focal plane can be determined independently to enable image capture settings to be tailored based upon factors including (but not limited to) a specific color channel and/or a specific portion of the scene dynamic range. The sensor elements utilized in the focal planes can be individual light sensing elements such as, but not limited to, traditional CIS (CMOS Image Sensor) pixels, CCD (charge-coupled device) pixels, high dynamic range sensor elements, multispectral sensor elements and/or any other structure configured to generate an electrical signal indicative of light incident on the structure. In many embodiments, the sensor elements of each focal plane have similar physical properties and receive light via the same optical channel and color filter (where present). In other embodiments, the sensor elements have different characteristics and, in many instances, the characteristics of the sensor elements are related to the color filter applied to each sensor element.

In several embodiments, color filters in individual cameras can be used to pattern the camera module with π filter groups as further discussed in U.S. Provisional Patent Application No. 61/641,165 entitled "Camera Modules Patterned with pi Filter Groups" filed May 1, 2012, the disclosure of which is incorporated by reference herein in its entirety. These cameras can be used to capture data with respect to different colors, or a specific portion of the spectrum. In contrast to applying color filters to the pixels of the camera, color filters in many embodiments of the invention are included in the lens stack. Any of a variety of color filter configurations can be utilized including the configuration in FIG. **1**C including

eight Green cameras, four Blue cameras, and four Red cameras, where the cameras are more evenly distributed around the center of the camera. For example, a Green color camera can include a lens stack with a Green light filter that allows Green light to pass through the optical channel. In many embodiments, the pixels in each focal plane are the same and the light information captured by the pixels is differentiated by the color filters in the corresponding lens stack for each filter plane. Although a specific construction of a camera module with an optic array including color filters in the lens stacks is described above, camera modules including it filter groups can be implemented in a variety of ways including (but not limited to) by applying color filters to the pixels of the focal planes of the camera module similar to the manner in which color filters are applied to the pixels of a conventional color camera. In several embodiments, at least one of the cameras in the camera module can include uniform color filters applied to the pixels in its focal plane. In many embodiments, a Bayer filter pattern is applied to the pixels of one of the cameras in a camera module. In a number of embodiments, camera modules are constructed in which color filters are utilized in both the lens stacks and on the pixels of the imager array.

Color Filters in Individual Cameras

In many embodiments, camera modules of an array camera are patterned with one or more π filter groups. The term patterned here refers to the use of specific color filters in individual cameras within the camera module so that the cameras form a pattern of color channels within the array camera. The term color channel or color camera can be used to refer to a camera that captures image data within a specific portion of the spectrum and is not necessarily limited to image data with respect to a specific color. The term Bayer camera can be used to refer to a camera that captures image data using the Bayer filter pattern on the image plane. In many embodiments, a color channel can include a camera that captures infrared light, ultraviolet light, extended color and any other portion of the visible spectrum appropriate to a specific application. The term π filter group refers to a 3×3 group of cameras including a central camera and color cameras distributed around the central camera to minimize occlusion zones. The central camera of a π filter group can be used as a reference camera when synthesizing an image using image data captured by an imager array. A camera is a reference camera when its viewpoint is used as the viewpoint of the synthesized image. The central camera of a π filter group is surrounded by color cameras in a way that minimizes occlusion zones for each color camera when the central camera is used as a reference camera. Occlusion zones are areas surrounding foreground objects not visible to cameras that are spatially offset from the reference camera due to the effects of parallax. In several embodiments, the central camera is a green camera while in other embodiments the central camera captures image data from any appropriate portion of the spectrum. In a number of embodiments, the central camera is a Bayer camera (i.e. a camera that utilizes a Bayer filter pattern to capture a color image). In many embodiments, a π filter group is a 3×3 array of cameras with a green color camera at each corner and a green color camera at the center which can serve as the reference camera with a symmetrical distribution of red and blue cameras around the central green camera. The symmetrical distribution can include arrangements where either red color cameras are directly above and below the center green reference camera with blue color cameras directly to the left and right, or blue color cameras directly above and below the green center reference camera with red color cameras directly to the left and right.

Camera modules of dimensions greater than a 3×3 array of cameras can be patterned with π filter groups in accordance with many embodiments of the invention. In many embodiments, patterning a camera module with π filter groups enables an efficient distribution of cameras around a reference camera that reduces occlusion zones. In several embodiments, patterns of π filter groups can overlap with each other such that two overlapping π filter groups on a camera module share common cameras. When overlapping π filter groups do not span all of the cameras in the camera module, cameras that are not part of a π filter group can be assigned a color to minimize occlusion zones in the resulting camera array.

In several embodiments, patterning a camera module with π filter groups can result in reference cameras that are not in the center of the camera module. Additionally, color cameras surrounding the reference camera need not be uniformly distributed but need only be distributed in a way to minimize occlusion zones of each color from the perspective of the reference camera. Utilization of a reference camera in a π filter group to synthesize an image from captured image data can be significantly less computationally intensive than synthesizing an image using the same image data from a virtual viewpoint.

High quality images or video can be captured by an array camera including a camera module patterned with π filter groups utilizing a subset of cameras within the camera module (i.e. not requiring that all cameras on a camera module be utilized). Similar techniques can also be used for efficient generation of stereoscopic 3D images utilizing image data captured by subsets of the cameras within the camera module.

Patterning camera modules with π filter groups also enables robust fault tolerance in camera modules with multiple π filter groups as multiple possible reference cameras can be utilized if a reference camera begins to perform sub optimally. Patterning camera modules with π filter groups also allows for yield improvement in manufacturing camera modules as the impact of a defective focal plane on a focal plane array can be minimized by simply changing the pattern of the color lens stacks in an optic array.

In several embodiments, color filters in individual cameras can be used to pattern the camera module with π filter groups. These cameras can be used to capture data with respect to different colors, or a specific portion of the spectrum. In contrast to applying color filters to the pixels of the camera, color filters in many embodiments of the invention are included in the lens stack. For example, a green color camera can include a lens stack with a green light filter that allows green light to pass through the optical channel. In many embodiments, the pixels in each focal plane are the same and the light information captured by the pixels is differentiated by the color filters in the corresponding lens stack for each filter plane. Although a specific construction of a camera module with an optic array including color filters in the lens stacks is described above, camera modules including π filter groups can be implemented in a variety of ways including (but not limited to) by applying color filters to the pixels of the focal planes of the camera module similar to the manner in which color filters are applied to the pixels of a conventional color camera. In several embodiments, at least one of the cameras in the camera module can include uniform color filters applied to the pixels in its focal plane. In many embodiments, a Bayer filter pattern is applied to the pixels of one of the cameras in a camera module. In a number of embodiments, camera modules are constructed in which color filters are utilized in both the lens stacks and on the pixels of the imager array.

Patterning with π Filter Groups

Camera modules can be patterned with π filter groups in accordance with embodiments of the invention. In several embodiments, π filter groups utilized as part of a camera module can each include a central camera that can function as a reference camera surrounded by color cameras in a way that reduces occlusion zones for each color. In certain embodiments, the camera module is arranged in a rectangular format utilizing the RGB color model where a reference camera is a green camera surrounded by red, green and blue cameras. In several embodiments, a number of green cameras that is twice the number of red cameras and twice the number of blue cameras surround the reference camera. However, any set of colors from any color model can be utilized to detect a useful range of colors in addition to the RGB color model, such as the cyan, magenta, yellow and key (CMYK) color model or red, yellow and blue (RYB) color model.

In several embodiments, two π filter groups can be utilized in the patterning of a camera module when the RGB color model is used. One π filter group is illustrated in FIG. **21**A and the other π filter group is illustrated FIG. **21**B. Either of these π filter groups can be used to pattern any camera module with dimensions greater than a 3×3 array of cameras.

In embodiments with a 3×3 camera module, patterning of the camera module with π filter group includes only a single π filter group. A π filter group on a 3×3 camera module in accordance with an embodiment of the invention is illustrated in FIG. **21**A. The π filter group **2100** includes a green camera at each corner, a green reference camera in the center notated within a box **2104**, blue cameras above and below the reference camera, and red cameras to the left and right sides of the reference camera. In this configuration, the number of green cameras surrounding the central reference camera is twice the number of red cameras and twice the number of blue cameras. An alternative to the π filter group described in FIG. **21**A is illustrated in FIG. **21**B in accordance with an embodiment of the invention. This π filter group also includes green cameras at the corners with a green reference camera **2152** at the center, as denoted with a box. However, unlike FIG. **21**A, the red cameras shown in FIG. **21**B are above and below, and the blue cameras are to the left and right side of the reference camera. As with the π filter group shown in FIG. **21**A, the π filter group in FIG. **21**B includes a central reference camera surrounded by a number of green cameras that is twice the number of red cameras and twice the number of blue cameras. As discussed above, the reference camera need not be a green camera. In several embodiments, the configurations in FIGS. **21**A and **21**B can be modified to include a central camera that employs a Bayer color filter. In other embodiments, the central camera is an infrared camera, an extended color camera and/or any other type of camera appropriate to a specific application. In further embodiments, any of a variety of color cameras can be distributed around the reference camera in a manner that reduces occlusion zones with respect to each color channel.

Any camera module with dimensions at and above 3×3 cameras can be patterned with one or more π filter groups, where cameras not within a π filter group are assigned a color that reduces or minimizes the likelihood of occlusion zones within the camera module given color filter assignments of the π filter groups. A 4×4 camera module patterned with two π filter groups in accordance with an embodiment of the invention is illustrated in FIG. **22**. The camera module **2200** includes a first π filter group **2202** of nine cameras centered on a reference green camera **2204**. A second π filter group **2210** is diagonally located one camera shift to the lower right of the first π filter group. The second π filter group shares the four center cameras **2212** of the camera module **2200** with the first π filter group. However, the cameras serve different roles (i.e. different cameras act as reference cameras in the two π filter groups). As illustrated in FIG. **22**, the two cameras at the corners **2206** and **2208** of the camera module are not included in the two π filter groups, **2202** and **2210**. The color filters utilized within these cameras are determined based upon minimization of occlusion zones given the color filter assignments of the cameras that are part of the two π filter groups, **2202** and **2210**. Due to the patterning of the π filter groups, there is an even distribution of blue color cameras around the reference camera, but there is no red color camera above the reference camera. Therefore, selecting the upper right corner camera **2206** to be red provides red image data from a viewpoint above the reference camera and the likelihood of occlusion zones above and to the right of the foreground images in a scene for the reference camera **2204** and the center camera of the second π filter group is minimized. Similarly, selecting the lower left corner camera **2208** to be blue provides blue image data from a viewpoint to the left of the reference camera and the likelihood of occlusion zones below and to the left of the foreground images in a scene for the reference camera **2204** and the center camera of the second π filter group is minimized. Thereby, a camera module with dimensions greater than 3×3 can be patterned with π filter groups with colors assigned to cameras not included in any π filter group to reduce and/or minimize occlusion zones as discussed above. Although specific π filter groups are discussed above, any of a variety of π filter groups can pattern a camera module in accordance with many different embodiments of the invention.

Multiple Reference Camera Options with Equivalent Performance

The use of multiple π filter groups to pattern a camera module in accordance with embodiments of the invention enables multiple cameras to be used as the reference camera with equivalent performance. A 4×4 camera module with two π filter groups in accordance with an embodiment of the invention is illustrated in FIG. **23**. The camera module **2300** includes two π filter groups **2302**, **2306** where the central camera of each π filter group **2304**, **2308** can act as a reference camera. Irrespective of the reference camera that is selected, the distribution of cameras around the reference camera is equivalent due to the use of π filter groups. Thereby, if a camera module **2300** detects a defect with the a reference camera **2304**, the camera module **2300** can switch to using the camera at the center of another π filter group as a reference camera **2308** to avoid the defects of the first reference camera **2304**. Furthermore, patterning with π filter groups does not require that the reference camera or a virtual viewpoint be at the center of a camera module but rather that the reference camera is surrounded by color cameras in a way that reduces occlusion zones for each color. Although a specific camera module is discussed above, camera modules of any number of different dimensions can be utilized to create multiple reference camera options in accordance with embodiments of the invention.

Capturing Images Using a Subset of Cameras

Array cameras with camera modules patterned with π filter groups can utilize less than all of the available cameras in operation in accordance with many embodiments of the invention. In several embodiments, using fewer cameras can minimize the computational complexity of generating an image using an array camera and can reduce the power consumption of the array camera. Reducing the number of cameras used to capture image data can be useful for applications such as video, where frames of video can be synthesized

using less than all of the image data that can be captured by a camera module. In a number of embodiments, a single π filter group can be utilized to capture an image. In many embodiments, image data captured by a single π filter group is utilized to capture a preview image prior to capturing image data with a larger number of cameras. In several embodiments, the cameras in a single IF filter group capture video image data. Depending upon the requirements of a specific application, image data can be captured using additional cameras to increase resolution and/or provide additional color information and reduce occlusions.

A π filter group within a camera module that is utilized to capture image data that can be utilized to synthesize an image is illustrated in FIG. **24**. In the illustrated embodiments, the reference camera is boxed and utilized cameras are encompassed in a dotted line. The camera module **2400** includes a π filter group of cameras generating image data G_1-G_2, G_5-G_6, B_1-B_2 and R_2-R_3 with reference camera G_3. FIG. **24** illustrates how the cameras in a π filter group can be utilized to capture images. Image data can be acquired using additional cameras for increased resolution and to provide additional color information in occlusion zones. Accordingly, any number and arrangement of cameras can be utilized to capture image data using a camera module in accordance with many different embodiments of the invention.

Although specific array cameras and imager arrays are discussed above, many different array cameras can be utilized to capture image data and synthesize images in accordance with embodiments of the invention. Systems and methods for detecting and correcting parallax in image data captured by an array camera in accordance with embodiments of the invention are discussed below.

Determining Parallax/Disparity

In a number of embodiments, the individual cameras in the array camera used to capture the light field have similar fields of view, fixed apertures, and focal lengths. As a result, the cameras tend to have very similar depth of field. Parallax in a two camera system is illustrated in FIG. **2**. The two cameras **200**, **202**, include a lens stack **204** and a focal plane **206**. Each camera has a back focal length f, and the two cameras are separated by the baseline distance of 2 h. The field of view of both cameras encompasses a scene including a foreground object **208** and a background object **210**. The scene from the viewpoint of the first camera **200** is illustrated in FIG. **3**A. In the image **300** captured by the first camera, the foreground object **208** appears located slightly to the right of the background object **210**. The scene from the viewpoint of the second camera **202** is illustrated in FIG. **3**B. In the image **302** captured by the second camera, the foreground object **208** appears shifted to the left hand side of the background object **210**. The disparity introduced by the different fields of view of the two cameras **200**, **202**, is equal to the difference between the location of the foreground object **208** in the image captured by the first camera (indicated in the image captured by the second camera by ghost lines **304**) and its location in the image captured by the second camera. As is discussed further below, the distance from the two cameras to the foreground object can be obtained by determining the disparity of the foreground object in the two captured images.

Referring again to FIG. **2**, the point (x_0, Y_0, z_0) on the foreground object will appear on the focal plane of each camera at an offset from the camera's optical axis. The offset of the point on the focal plane of the first camera **200** relative to its optical axis **212** is shown as $-u_L$. The offset of the point on the focal plane of the second camera **202** relative to its optical axis **214** is shown as U_R. Using similar triangles, the offset between the images captured by the two cameras can be observed as follows:

$$\frac{h - x_o}{z_o} = \frac{-u_L}{f}$$

$$\frac{h + x_o}{z_o} = \frac{u_R}{f}$$

Combining the two equations yields the disparity (or parallax) between the two cameras as:

$$\Delta_{parallax} = u_R - u_L = \frac{2hf}{z_o}$$

From the above equation, it can be seen that disparity between images captured by the cameras is along a vector in the direction of the baseline of the two cameras, which can be referred to as the epipolar line between the two cameras. Furthermore, the magnitude of the disparity is directly proportional to the baseline separation of the two cameras and the back focal length of the cameras and is inversely proportional to the distance from the camera to an object appearing in the scene.

Occlusions in Array Cameras

When multiple images of a scene are captured from different perspectives and the scene includes foreground objects, the disparity in the location of the foreground object in each of the images results in portions of the scene behind the foreground object being visible in some but not all of the images. A pixel that captures image data concerning a portion of a scene, which is not visible in images captured of the scene from other viewpoints, can be referred to as an occluded pixel. Referring again to FIGS. **3**A and **3**B, when the viewpoint of the second camera is selected as a reference viewpoint the pixels contained within the ghost lines **304** in the image **302** can be considered to be occluded pixels (i.e. the pixels capture image data from a portion of the scene that is visible in the image **302** captured by the second camera **202** and is not visible in the image **300** captured by the first camera **200**). The pixels contained in the ghost line **306** in the first image can be considered to be revealed pixels (i.e. pixels that are not visible in the reference viewpoint, but that are revealed by shifting to an alternate viewpoint). In the second image, the pixels of the foreground object **208** can be referred to as occluding pixels as they capture portions of the scene that occlude the pixels contained within the ghost lines **304** in the image **302**. Due to the occlusion of the pixels contained within the ghost lines **304** in the second image **302**, the distance from the camera to portions of the scene visible within the ghost lines **304** cannot be determined from the two images as there are no corresponding pixels in the image **300** shown in FIG. **3**A.

As is discussed further below, increasing the number of cameras capturing images of a scene from different viewpoints in complementary occlusion zones around the reference viewpoint increases the likelihood that every portion of the scene visible from the reference viewpoint is also visible from the viewpoint of at least one of the other cameras. When the array camera uses different cameras to capture different wavelengths of light (e.g. RGB), distributing at least one camera that captures each wavelength of light in the quad-

rants surrounding a reference viewpoint can significantly decrease the likelihood that a portion of the scene visible from the reference viewpoint will be occluded in every other image captured within a specific color channel. The distribution of color filters in array cameras to reduce the likelihood of occlusions in accordance with embodiments of the invention is discussed further in U.S. Provisional Patent Application Ser. No. 61/641,164 entitled "Camera Modules Patterned with π Filter Groups", to Nisenzon et al., filed May 1, 2012, the disclosure of which is incorporated herein by reference in its entirety.

Using Disparity to Generate Depth Maps in Array Cameras

Array cameras in accordance with many embodiments of the invention use disparity observed in images captured by the array cameras to generate a depth map. A depth map is typically regarded as being a layer of metadata concerning an image that describes the distance from the camera to specific pixels or groups of pixels within the image (depending upon the resolution of the depth map relative to the resolution of the original input images). Array cameras in accordance with a number of embodiments of the invention use depth maps for a variety of purposes including (but not limited to) generating scene dependent geometric shifts during the synthesis of a high resolution image and/or performing dynamic refocusing of a synthesized image.

Based upon the discussion of disparity above, the process of determining the depth of a portion of a scene based upon pixel disparity is theoretically straightforward. When the viewpoint of a specific camera in the array camera is chosen as a reference viewpoint, the distance to a portion of the scene visible from the reference viewpoint can be determined using the disparity between the corresponding pixels in some or all of the images captured by the camera array. In the absence of occlusions, a pixel corresponding to a pixel in the image captured from the reference viewpoint will be located in each non-reference or alternate view image along an epipolar line (i.e. a line parallel to the baseline vector between the two cameras). The distance along the epipolar line of the disparity corresponds to the distance between the camera and the portion of the scene captured by the pixels. Therefore, by comparing the pixels in the captured images that are expected to correspond at a specific depth, a search can be conducted for the depth that yields the pixels having the highest degree of similarity. The depth at which the corresponding pixels in the captured images have the highest degree of similarity can be selected as the most likely distance between the camera and the portion of the scene captured by the pixel. As is discussed below, similarity can be determined with respect to corresponding pixels within a single spectral channel, within multiple spectral channels, and/or across spectral channels as appropriate to the requirements of specific applications in accordance with embodiments of the invention.

Many challenges exist, however, in determining an accurate depth map using the method outlined above. In several embodiments, the cameras in an array camera are similar but not the same. Therefore, image characteristics including (but not limited to) optical characteristics, different sensor characteristics (such as variations in sensor response due to offsets, different transmission or gain responses, non-linear characteristics of pixel response), noise in the captured images, and/or warps or distortions related to manufacturing tolerances related to the assembly process can vary between the images reducing the similarity of corresponding pixels in different images. In addition, super-resolution processes rely on sampling diversity in the images captured by an imager array in order to synthesize higher resolution images. However, increasing sampling diversity can also involve decreasing similarity between corresponding pixels in captured images in a light field. Given that the process for determining depth outlined above relies upon the similarity of pixels, the presence of photometric differences and sampling diversity between the captured images can reduce the accuracy with which a depth map can be determined.

The generation of a depth map is further complicated by occlusions. As discussed above, an occlusion occurs when a pixel that is visible from the reference viewpoint is not visible in one or more of the captured images. The effect of an occlusion is that at the correct depth, the pixel location that would otherwise be occupied by a corresponding pixel is occupied by a pixel sampling another portion of the scene (typically an object closer to the camera). The occluding pixel is often very different to the occluded pixel. Therefore, a comparison of the similarity of the pixels at the correct depth is less likely to result in a significantly higher degree of similarity than at other depths. Effectively, the occluding pixel acts as a strong outlier masking the similarity of those pixels, which in fact correspond at the correct depth. Accordingly, the presence of occlusions can introduce a strong source of error into a depth map.

Processes for generating depth maps in accordance with many embodiments of the invention attempt to minimize sources of error that can be introduced into a depth map by sources including (but not limited to) those outlined above. A general process for generating a depth map in accordance with an embodiment of the invention is illustrated in FIG. **4**. The process **400** involves capturing (**402**) a light field using an array camera. In a number of embodiments, a reference viewpoint is selected (**404**). In many embodiments, the reference viewpoint is predetermined. In several embodiments, the reference viewpoint can be determined based upon the captured light field or a specific operation requested by a user of the array camera (e.g. generation of a stereoscopic 3D image pair). Prior to determining a depth map, the raw image data is normalized (**406**) to increase the similarity of corresponding pixels in the captured images. In many embodiments, normalization involves utilizing calibration information to correct for variations in the images captured by the cameras including (but not limited to) photometric variations and scene-independent geometric distortions introduced by each camera's lens stack. In several embodiments, the normalization of the raw image data also involves pre-filtering to reduce the effects of aliasing and noise on the similarity of corresponding pixels in the images, and/or rectification of the image data to simplify the geometry of the parallax search. The filter can be a Gaussian filter or an edge-preserving filter, a fixed-coefficient filter (box) and/or any other appropriate filter. In a number of embodiments, normalization also includes resampling the captured images to increase the similarity of corresponding pixels in the captured images by correcting for geometric lens distortion, for example. Processes performed during the normalization or raw image data in accordance with embodiments of the invention are discussed further below.

An initial depth map is determined (**408**) for the pixels of an image captured from the reference viewpoint. The initial depth map is used to determine (**410**) likely occlusion zones and the depths of pixels in the occlusion zones are updated (**412**) by determining the depths of the pixels in occlusion zones using images in which a corresponding pixel is visible. As is discussed further below, depth estimates can be updated using competing subsets of images corresponding to different visibility patterns encountered in real world scenes. Although a specific sequence is shown in FIG. **4**, in many embodiments occlusion zones are detected at the same time the initial depth map is generated.

A normalization process involving resampling the raw image data to reduce scene-independent geometric differences can reduce errors by correcting linear and/or non-linear lens distortion which might otherwise compromise the ability to match corresponding pixels in each of the captured images. In addition, updating the depth map in occlusion zones with depth measurements that exclude occluded pixels further reduces sources of error in the resulting depth map. Although a general process for generating a depth map is illustrated in FIG. **4**, variations and alternatives to the illustrated processes for generating depth maps can be utilized in accordance with embodiments of the invention. Processes for calibrating raw image data, determining initial depth maps, and for updating depth maps to account for occlusions in accordance with embodiments of the invention are discussed further below.

Increasing Similarity of Corresponding Pixels in Captured Image Data

The greater the similarity between the images captured by each of the cameras in an array camera, the higher the likelihood that a measurement of corresponding pixels in the images at different hypothesized depths will result in highest similarity being detected at the correct depth. As is disclosed in U.S. patent application Ser. No. 12/935,504 (incorporated by reference above) the images captured by cameras in an array camera typically differ in a number of ways including (but not limited to) variations in the optics from one camera to another can introduce photometric differences, aliasing, noise, and scene-independent geometric distortions. Photometric differences and scene-independent geometric distortions can be corrected through filtering and calibration. Photometric calibration data used to perform photometric normalization and scene-independent geometric corrections that compensate for scene-independent geometric distortions can be generated using an off line calibration process and/or a subsequent recalibration process. The photometric calibration data can be provided to a photometric normalization module or process that can perform any of a variety of photometric adjustments to the images captured by an array camera including (but not limited to) pre-filtering to reduce the effects of aliasing and noise, Black Level calculation and adjustments, vignetting correction, and lateral color correction. In several embodiments, the photometric normalization module also performs temperature normalization. The scene-independent geometric corrections determined using a calibration process can also be applied to the captured images to increase the correspondence between the images. When the captured images are used to synthesize a higher resolution image using super-resolution processing, the scene-independent geometric corrections applied to the images are typically determined at a sub-pixel resolution. Accordingly, the scene-independent geometric corrections are typically determined with a higher degree of precision than the corrections utilized during registration in conventional stereoscopic 3D imaging. In many embodiments, the scene-independent geometric corrections also involve rectification to account for distortion and rotation of the lenses of the array camera relative to the focal planes so that the epipolar lines of the non-reference images are easily aligned with those of the image captured from the reference viewpoint. By normalizing geometrically in this way, the searches performed to determine the depths of corresponding pixels can be simplified to be searches along straight lines in various cameras, and the precision of depth measurements can be improved.

Systems and methods for calibrating array cameras to generate a set of scene-independent geometric corrections and photometric corrections that can be applied to images captured by an array camera in accordance with embodiments of the invention are described in U.S. Patent Application Ser. No. 61/780,748, entitled "Systems and Methods for Calibration of an Array Camera" to Mullis, Jr., filed Mar. 13, 2013, the disclosure of which is incorporated by reference in its entirety.

In a number of embodiments, the correspondence of the pixels in the captured images is increased by resampling the images to detect objects to sub-pixel precision shifts in the fields of view of the cameras in the array camera.

A process for applying corrections to images captured by an array camera to increase the correspondence between the captured images in accordance with embodiments of the invention is illustrated in FIG. **5**. The process **500** includes photometrically normalizing the captured images (**502**), applying scene-independent geometric corrections (**504**) to the normalized images. In some embodiments, an additional rectification process (**505**) is needed to ensure that all cameras are co-planar and parallax search can be reduced to epipolar lines only. The processes shown in FIG. **5** increase the correspondence between the resulting images. Therefore, searches for pixel correspondence between the images are more likely to result in accurate depth measurements.

Although specific processes for increasing the correspondence between images captured by an array camera(s) in accordance with embodiments of the invention are discussed above with respect to FIG. **5**, any of a variety of processes that increase the correspondence between the captured images can be utilized prior to generating a depth map in accordance with embodiments of the invention. Processes for generating depth maps in accordance with embodiments of the invention are discussed further below.

Generating a Depth Map

The process of generating a depth map involves utilizing disparity between images to estimate the depth of objects within a scene. As noted above, occlusions can impact the reliability of depth measurements obtained using cost functions in the manner outlined above. Typically such occlusions will manifest themselves as significant mismatches according to the similarity metric used to compare corresponding pixels (potentially masking the similarity of the visible pixels). However, many embodiments of the invention generate an initial depth map and then address any errors that may have been introduced into the creation of the initial depth map by occlusions. In several embodiments, the initial depth map is utilized to identify pixels in the image captured from a reference viewpoint that may be occluded in images captured by the array camera from other viewpoints. When an occlusion is detected, the depth information for the pixel in the image captured from the reference viewpoint can be updated by excluding pixels from the image in which the pixel is occluded from the similarity comparisons. In several embodiments, depth estimates impacted by occlusions can be updated using competing subsets of images corresponding to different visibility patterns encountered in real world scenes. In certain embodiments, the updated depth estimates can be utilized to identify corresponding pixels that are occluded and the depth estimation process iterated using the visibility information so that the impact of occlusions on the precision of the depth map can be reduced. In several embodiments, the process of generating updated depth estimates using subsets of images is sufficiently robust that the need to iteratively refine the depth map and visibility estimates can be reduced or eliminated.

A process for determining depth of pixels in an image captured from a reference viewpoint in accordance with an embodiment of the invention is illustrated in FIG. **6**. The process **600** includes determining (**602**) an initial depth map

using some or all of the images captured by an array camera. The visibility of each pixel in the image captured from the reference viewpoint in each of the captured images is then determined (**604**). Where a corresponding pixel location is occluded, the depth of the pixel in the image captured from the reference viewpoint can be recalculated (**606**) excluding the image in which the corresponding pixel location is occluded from the cost function. A decision (**608**) is made concerning whether to continue to iterate. As depth measurements in occlusion zones are refined, additional information is obtained concerning the visibility of pixels within the occlusion zones in each of the captured images. Therefore, repeating the recalculation of the depths of pixels in the occlusion zones as the visibility information is refined can iteratively improve the precision of the depth map. Any of a variety of termination conditions appropriate to a specific application can be utilized to determine when to terminate the iterative loop including (but not limited to) the completion of a predetermined number of iterations and/or the number of pixels for which updated depth information is determined in a specific pass through the iterative loop falling below a predetermined number. In several embodiments, a single iteration only is performed due to the exploitation of subsets of the set of images corresponding to real world visibility patterns to update depth estimates generated using mismatched pixels.

Once a finalized depth map is obtained, the visibility of each of the pixels in the captured images is determined (**610**) and the depth map and/or visibility information can be utilized for a variety of purposes including but not limited to the synthesis of a high resolution image using super-resolution processing.

The computational complexity of a process similar to the process illustrated in FIG. **7** depends on the number of images compared when performing depth determinations. The further a camera is from the reference viewpoint the larger the disparity that will be observed. In addition, the furthest cameras in the array encompass all the other cameras within their envelope. Typically larger magnitude shifts enable depth to be determined with greater precision. Therefore, using a camera that captures an image from a reference viewpoint and the cameras that are furthest from that camera to determine depth information can improve precision of the detected depth. In addition, using an aggregated cost originating from cameras with various baselines and directions can significantly improve reliability of depth estimates due to increased likelihood of reducing periodicity in matches. In the case of a 5×5 array (see FIG. **7**), the central Green camera (**700**) can be utilized to capture an image from a reference viewpoint and the image data captured by the central camera can be compared to image data captured by the Green cameras (**702**) located in the four corners of the array to determine depth. In other arrays, images captured by any of a variety of combinations of cameras can be utilized to determine depth in accordance with embodiments of the invention. As is discussed further below, selection of specific subsets of cameras can decrease the likelihood that a pixel in a reference image will be occluded in image data captured by other cameras in the subset.

Although a specific process for generating a depth map and/or visibility information in accordance with an embodiment of the invention is illustrated in FIG. **6**, any of a variety of processes can be utilized that involve determining an initial depth map and then refining the depth map by detecting occluded pixels and updating the depth measurements to exclude occluded pixels. Specific processes for determining depth and visibility of pixels in accordance with embodiments of the invention are discussed further below.

Determining an Initial Depth Map

Processes for determining the distance from an array camera to an object in a scene involve locating the depth at which corresponding pixels in images captured by the array camera have the highest degree of similarity. As discussed above, at a specific depth a pixel in an image captured from a reference viewpoint will shift a known distance along an epipolar line between the reference viewpoint and each of the cameras in the camera array. The pixel in the image captured from the reference viewpoint and the "shifted" pixels in the other images (i.e. the pixels in the images located in locations determined based upon the anticipated shift for a specific distance) are the corresponding pixels. When a hypothesized depth is incorrect, the corresponding pixels may exhibit very little similarity (although in some scenes incorrect depths have high degrees of similarity due to features such as periodic texture). When the hypothesized depth is correct, the corresponding pixels will ideally exhibit the highest degree of similarity of any of the hypothesized depths. When a depth map is used in super-resolution processing of a captured light field, a depth map can be determined with sufficient precision to enable detection of sub-pixel shifts. In super-resolution processing, it is the scene-dependent shifts that are utilized and not the depth directly. Therefore, the ability to detect depth corresponding to sub-pixel shift precision can significantly improve the performance of the super-resolution processing. The manner in which resampling of the pixels of the captured images can be utilized to determine depth with sub-pixel shift precision is discussed further below.

In many embodiments, a parallax search of a number of depths within a range in physical distance (e.g. 20 cm to infinity) is utilized to inform the disparities searched when performing depth estimation. The search range can be divided into a number of depth indices such that the parallax shifts between consecutive depth indices is constant in pixels for a particular image and is set based upon a minimum sub-pixel precisions as measured for the images captured by cameras in the array corresponding to the largest baselines with respect to the reference viewpoint (see for example FIG. **7**). This increases the likelihood of sufficient accuracy in the depth estimates for use as inputs to a super-resolution process. In other embodiments, consecutive depth indices need not correspond to constant pixel shifts and the depth search can adapt based upon the characteristics of the scene.

In several embodiments, a cost function is utilized to determine the similarity of corresponding pixels. The specific cost function that is used typically depends upon the configuration of the array camera, the number of images captured by the array camera, and the number of color channels utilized by the array camera. In a number of embodiments, the array camera includes a single color channel and/or a depth map is generated using cameras within a single color channel. Where image data from within a single color channel is utilized to generate the depth map, a cost function can be utilized that measures the variance of the corresponding pixels. In several embodiments, sums of L1 norms, L2 norms, or some other metrics can be used. For example, the aggregation of similarity metrics with respect to a target (typically reference but non-reference may also be used). The smaller the variance, the greater the similarity between the pixels.

Image data from multiple spectral channels can also be utilized to generate a depth map. In several embodiments, the depth at a given pixel location is estimated by looking at the similarity of corresponding pixels from images within each of the spectral channels. In a number of embodiments, the pro-

cess of determining the depth at a given pixel location also involves using information concerning the similarity of corresponding pixels from images across different spectral channels. Cost functions that can be utilized when generating a depth map using image data captured using multiple color channels include (but are not limited to) L1 norms, L2 norms, or a combination of L1 and L2 norms, of the combinations of image data from the different color channels and/or the variance/standard deviation of corresponding pixels within multiple individual color channels. In other embodiments, truncated versions of the L1 and L2 norms and/or any block-based similarity measure based on rank, census, correlation, and/or any other appropriate metric such as those practiced in multiview stereo disparity detection techniques can be utilized.

As is discussed further below, many embodiments of the invention utilize subsets of cameras including cameras from multiple color channels grouped based upon characteristics of natural scenes when determining the depth of a pixel location in an image form a reference viewpoint to decrease the likelihood that a given pixel location is occluded in the alternate view images captured by the other cameras in the subset of cameras. Where an array camera utilizes a Bayer filter in the camera that captures an image from the reference viewpoint, then a variety of cost functions can be utilized to determine pixel similarity including (but not limited to) cost functions that measure the combination of Red variance, Green variance, and Blue variance. In addition, different cost functions can be applied to the pixels in different regions of an image. In several embodiments, a depth map is generated from image data captured by a central Green camera and a cluster of Red, Blue and Green cameras in each of the four corners of a camera array using this technique (see for example FIG. **7**).

A process for determining the depth of a pixel using images captured by an array camera in accordance with an embodiment of the invention is illustrated in FIG. **8**. The process **800** includes selecting (**802**) an initial hypothesized depth or distance d for a selected pixel from an image captured from a reference viewpoint. Based upon the location of the pixel within the reference image and information concerning the baseline between the reference viewpoint and the viewpoints of the other cameras used to perform the depth measurement, the corresponding pixel locations in each of the captured images at the hypothesized depth d are determined (**804**). In many embodiments, the input images to the parallax detection process are not geometrically corrected, and the geometric correction is applied on-the-fly by adding a vector offset to the parallax shift during the search to identify corresponding pixels at a given depth d. In other embodiments, the geometric correction is applied to the images before the search commences during a normalization process and no geometric correction vector must be added during the search when calculating pixel correspondences (i.e. the geometric corrections are pre-calculated). In the latter case, the pre-correction of geometric distortion can make the algorithm significantly more efficient on parallel processors such as SIMD and GPUs.

As noted above, occlusions can introduce errors into depth estimates. When occlusion/visibility information is available, occluded pixels can be disregarded (**806**) as part of the depth measurement. When information concerning the visibility of pixels is not available (e.g. during the generation of an initial depth map and/or during the generation of a depth estimate using a subset of images), the similarity of all of the pixels in the corresponding pixel locations is used to determine depth. As is discussed below with reference to FIGS. **8A-8I**, initial depth searches can be performed with respect to image data captured from subsets of images captured by the array camera to identify a specific subset of cameras in which a given pixel in the reference image is visible.

When the corresponding pixels have been identified, the similarity of the corresponding pixels can be measured (**808**). In many embodiments, the similarity of the pixels is determined using a cost function. The specific cost function utilized depends upon the pixel information that is compared. As noted above, in one embodiment, when pixels from a single color channel are compared the cost function can consider L1 norms, L2 norms, and/or the variance of corresponding pixels. When pixels from multiple color channels are compared, more complex cost functions can be utilized including (but not limited to) cost functions that incorporate the L1 and/or L2 norms of the image data from multiple color channels and/or the variance/standard deviation of corresponding pixels within multiple individual color channels. In other embodiments, truncated versions of the L1 and L2 norms and/or any block-based similarity measure based on rank, census, correlation, and/or any other appropriate metric such as those practiced in multiview stereo disparity detection techniques can be utilized. In several embodiments, the process of determining similarity utilizing a cost function involves spatially filtering the calculated costs using a filter such as (but not limited to) a fixed-coefficient filter (such as a Gaussian filter), or in an alternative embodiment, an edge-preserving filter. In the latter embodiment, filtering with an edge-preserving filter in this way is a form of adaptive support that utilizes information from photometrically similar neighboring pixels to improve the depth estimates. Without the filtering the depth measurements are pixel-wise and are noisier than if they are filtered. Smoothing the cost function using adaptive support can prevent the generation of incorrect depths. In a number of embodiments, the calculated costs are spatially filtered using a bilateral filter, where the bilateral filter weights are determined from the reference image but, in contrast to a normal bilateral filter, the resulting filter weights applied to the calculated costs and not the reference image. In this way the reference image data can be used as a guide to improve the denoising of the cost estimates. In a number of embodiments, a box filter and/or any other filter appropriate to the requirements of a specific application can be utilized.

The calculation of the cost function for corresponding pixels at different depths is repeated sweeping across a range of hypothesized depths (**812**) until the depth search is complete (**810**). The most likely depth can then be determined (**814**) as the hypothesized depth at which the (filtered) cost function indicates that the corresponding pixels have the highest level of similarity. In several embodiments, for a given depth computation early termination can occur if a single camera shows a very high mismatch. In this condition, the process can skip onto the next hypothesized depth since match at the current depth would be unacceptable. In many embodiments, the process of performing depth sampling (i.e. comparing pixels in alternate view images based upon the disparity at a specific depth) involves sampling depth uniformly in disparity space. Stated another way, depth samples can be taken at uniform pixel shifts along an epipolar line. In a number of embodiments, the search does not involve uniform sampling in disparity space. In several embodiments, the search exploits image characteristics to increase the efficiency of the search. In several embodiments, the search uses prior information about where objects are in the scene, such as from a coarser or lower spatial resolution depth map or reduced search resolution in disparity (e.g. from an image preview), to determine or restrict which depths are sampled in trying to form a higher resolution depth map. For example, a preview depth map may

be used to determine that there are no objects beyond a particular distance, in which case for the depth search, no depth samples would be allocated beyond that distance.

Many images exhibit regions of similar color, therefore, the search for the most likely hypothesized depth can be performed intelligently by selecting a first set of hypothesized depths that are more coarsely distributed across the range of possible hypothesized depths and then locating the depth among these that exhibits the highest degree of similarity. A second search can then be performed to refine within a more granular range of depths around the depth that exhibited the highest degree of similarity in the first set of depths. In the event that the more granular search fails and the best pixel found is not from a region exhibiting similar color, a full search can be performed across the entire range of depths at more precise intervals than in the original first coarse search. However, if a satisfactory match is found in the second search, the depth that exhibits the highest level of similarity within the second search can be used as the most likely hypothesized depth.

In many embodiments, searches for the most likely depth of a pixel are performed utilizing depth information determined for adjacent pixels. In several embodiments, the search is performed by searching around the depth of one or more adjacent pixels, by searching around a depth determined based on the depths of adjacent pixels (e.g. based on the average depth of adjacent pixels or based on linear interpolations of pairs adjacent pixels) and/or by searching around a previously identified depth (e.g. a depth determined with respect to a preview image and/or a previous frame in a video sequence). Searching in this way can also simplify the application of spatial filters when determining depth (see discussion below). In other embodiments, any of a variety of techniques can be utilized to reduce the computational complexity of locating the most likely depth of the pixels in an image.

Although specific processes for determining the depth of a pixel in an image captured from a reference viewpoint are discussed above with respect to FIG. **8**, any of a variety of processes can be utilized to determine the depth of a pixel including process that determine the depth of a pixel from a virtual viewpoint based upon a plurality of images captured by an array camera. Processes similar to the process illustrated in FIG. **8** can be utilized to generate an initial depth map and then to refine the depth map by ignoring images in which a corresponding pixel to a pixel location in an image from the reference viewpoint is occluded. Processes for determining pixel correspondence using adaptive support in accordance with embodiments of the invention are discussed further below.

Determining Pixel Correspondence in the Presence of Occlusions

Wherever there is a depth transition or discontinuity in the reference viewpoint, pixels adjacent the depth transition are likely to be occluded in at least one of the images captured by the array camera. Specifically, the pixels adjacent to the transition that are further in distance from the camera are likely to be occluded by the pixels adjacent the camera that are closer to the camera. Ideally, a depth map is determined using an aggregated cost function CV(x, y, d) for each visible camera i in the array that excludes occluded pixels as follows:

$$CV(x,y,d)=\sum_i \frac{Cost^{i,Ref}(x,y,d)\times V^{i,Ref}(x,y)}{\text{number of visible cameras at } (x,y)}$$

where $Cost^{i,Ref}(x, y, d)$ is a similarity measure (i.e. the cost function),

d is depth of pixel (x, y), and

$V^{i,Ref}(X, y)$ is the visibility of pixel (x, y) and initially $V^{i,Ref}(x, y)=1$ for all cameras.

In a number of embodiments, the individual costs $Cost^{i,Ref}(x,y,d)$ are computed based on each disparity hypothesis d for each pixel (x, y) for cameras i, Ref as follows:

$$Cost^{i,Ref}(x,y,d)=S\{I^i(x,y,d),I^{Ref}(x,y,d)\}$$

where S is the similarity measure (for example), and

I^i is the calibrated image i after geometric calibration.

In several embodiments, the process of generating an aggregated cost can involve use of images to which the scene-dependent geometric shifts corresponding to a specific hypothesized or candidate depth are applied to all pixels in the image. In this way, a shifted image can be generated for each candidate depth searched. Using the shifted images, an aggregated cost at each depth for a specific pixel location (x, y) in an image from the reference viewpoint can be generated in the manner outlined above utilizing the similarity between the shifted images and the reference image. In addition, the aggregated cost can consider the similarity of the shifted images at the candidate depth as follows:

$$CV(x,y,d)=\sum_{k\in K}\frac{(x,y)Cost^{k,Ref}(x,y,d)\times V^{k,ref}}{\text{number of cameras in } K}+\sum_{i,j\in L}\frac{Cost^{i,j}(x,y,d)\times V^{i,Ref}(x,y)\times V^{j,Ref}(x,y)}{\text{number of pairs of cameras in } L}$$

Where K is a set of cameras in the same spectral channel as the reference camera,

L is a set of pairs of cameras, where both cameras in each pair are in the same spectral channel (which can be a different spectral channel to the reference camera where the light field includes image data in multiple spectral channels),

$Cost^{k,Ref}(x, y, d)=S\{ImageRef(x, y), ShiftedImage^k(x, y, d)\}$, and

$Cost^{i,j}(x, y, d)=S\{ShiftedImage^i(x, y, d), ShiftedImage^j(x, y, d)\}$

In a number of embodiments, the sets K and L do not necessarily contain all cameras or pairs of cameras that satisfy the requirements in K and L. Furthermore, the cumulative cost function can also be constructed using a cost term in which the set of L includes arbitrarily large groups of cameras for which the cost of corresponding pixels is determined. In many embodiments, the similarity metric S is the L1 norm. In several embodiments, the similarity metric can be any of a number of well known similarity metrics including (but not limited to) the L2 norm, the variance or standard deviation of the corresponding pixels (particularly where L includes larger groups of cameras) window-based similarity metrics incorporating correlation, rank, census and/or any other measure appropriate to the requirements of a specific application. Although comparisons are discussed above in the context of shifted images, as can be readily appreciated comparisons can be performed by applying shifts to individual pixel locations

and comparing corresponding pixels at a hypothesized depth (as opposed to applying shifts to all pixels in an image and then comparing the shifted images).

In a number of embodiments, the cost function can also consider similarity between corresponding pixels across different spectral channels. In several embodiments, the similarity of neighborhoods of pixels in pixels from different spectral channels can be evaluated using any of a variety of metrics including (but not limited to) the cross-correlation of the pixels in the neighborhoods, the normalized cross-correlation between the pixels in the neighborhoods and/or any other metric for measuring the similarity of the relative values of two sets of pixels such as (but not limited to) entropy measures including measuring mutual information.

In several embodiments, different weightings can be applied to the similarity of corresponding pixels within a spectral channel containing a reference image and the reference image, the similarity of corresponding pixels within alternate view images in the same spectral channel, and/or the similarity of corresponding pixels within images in different spectral channels.

As discussed above, the aggregated cost function can be spatially filtered as follows:

$$\text{Filtered}CV(x,y,d)=\text{Filter}_{x_m,y_m\in N(x,y)}\{\text{Cost}(x_m,y_m,d)\}$$

where the Filter is applied in a neighborhood N(x, y) surrounding pixel location (x, y).

The filter can be a simple 3×3 or N×N box filter or some other filter including (but not limited to) a joint bilateral filter that uses the reference image as guidance, a fixed coefficient filter (such as a Gaussian filter, or a box filter), or any appropriate edge preserving filter. In several embodiments, the weighted aggregated cost function is as follows:

$$FilteredCV(x,\ y,\ d)=\frac{1}{Norm}\sum_{\substack{(x_1,y_1)\\ \in N(x,y)}} CV(x_1,\ y_1,\ d)\times wd(x,\ y,\ x_1,\ y_1)\times wr(I_{Ref}(x,\ y)-I_{Ref}(x_1,\ y_1))$$

where N(x, y) is the immediate neighborhood of the pixel (x, y), which can be square, circular, rectangular, or any other shape appropriate to the requirements of a specific application,

Norm is a normalization term,

$I_{Ref}(x, y)$ is the image data from the reference camera,

wd is a weighting function based on pixel distance, and

wr is a weighting function based on intensity difference.

In many embodiments, the filter is a bilateral filter and wd and wr are both Gaussian weighting functions.

Based upon the filtered aggregated cost function, a depth map can be computed by selecting the depth that minimizes the filtered cost at each pixel location in the depth map as follows:

$$D(x,y)=\text{argmin}_d\{\text{Filtered}CV(x,y,d)\}$$

When the aggregated cost function is filtered using an edge preserving filter in the manner outlined above, the likelihood that noise will result in the incorrect detection of occluded pixels is reduced. Instead of computing depths for individual pixels, an adaptive support window is used around each pixel to filter noise in a manner that preserves depth transitions. Utilizing a filter such as (but not limited to) a bilateral filter provides an adaptive window of support that adapts based upon the content. In many embodiments, a bilateral filter is used in which the reference image is used to define the spatial and range support for the bilateral filter (i.e. the parameters that define the size of the window of pixels that contribute to the aggregated cost function for a specific pixel). As a result, smoothing of the cost function of a pixel can be achieved using the calculated cost function of pixels that are part of the same surface. In other embodiments, filters such as (but not limited to) box filters are less computationally complex and provide sufficient filtering for the requirements of specific applications.

Determining Pixel Correspondence for Pixels in Multiple Spectral Channels

Array cameras in accordance with many embodiments of the invention include cameras in multiple spectral channels such as, but not limited to, Red, Green and Blue cameras. The cost metric CV (x, y, d) is described above in the context of a single spectral channel and multiple spectral channels. In the case of an array camera including Red, Green, and Blue cameras, the cost function can consider the similarity of pixels in the Green cameras, the similarity in pixels in the Red cameras, and the similarity of pixels in the Blue cameras at a particular depth. Where a camera in a specific color channel is chosen as the reference camera (e.g. a Green camera), pixels in the other channels (e.g. Red and Blue cameras) are difficult to directly compare to pixels in the reference image. However, the disparity at a particular depth can be determined and the intensity values of corresponding pixels in other color channels can be compared. Incorporating these additional comparisons into the depth estimate can improve depth estimates by utilizing information across all color channels. Various cost functions that can be utilized to perform depth estimation in array cameras that include Red, Green, and Blue cameras are discussed further below. As can be readily appreciated, however, the same cost functions can be utilized with respect to any set of spectral channels in accordance with embodiments of the invention.

In several embodiments, image data is captured using an array camera including Red, Green and Blue cameras and a Green camera is selected as a reference camera. A cost function can be utilized that considers pixel correspondence between pixels in a set of Green cameras, between pixels in a set of Red cameras, and between pixels in a set of Blue cameras when determining depth estimates. In several embodiments, the following cost function can be utilized:

$$\text{Cost}(x,\ y,\ d)=\gamma_G(x,\ y)\cdot Cost_G(x,\ y,\ d)+\gamma_R(x,\ y)\cdot Cost_R(x,\ y,\ d)+\gamma_B(x,\ y)\cdot Cost_B(x,\ y,\ d)$$

where $\text{Cost}_G(x, y, d)$ is the measure the similarity of pixels in locations within a set of Green cameras determined based upon the depth d and the location of the pixel (x, y) in the reference Green camera,

- $\text{Cost}_R(x, y, d)$ is the measure of the similarity of corresponding pixels in locations within a set of Red cameras determined based upon the depth d and the location of the pixel (x, y) in the reference Green camera,
- $\text{Cost}_B(x, y, d)$ is the measure of the similarity of corresponding pixels in locations within a set of Blue cameras determined based upon the depth d and the location of the pixel (x, y) in the reference Green camera, and
- γ_G, γ_R, and γ_B are weighting factors for the Green, Red and Blue cost functions respectively which may be constants for the entire reference viewpoint, or may vary spatially.

The spatial weighting may depend on the captured image data (for example using edge gradients), may correct or use known properties of the sensor (for example using a noise model prior for a given sensor to calculate SNR), as well as properties of the cost function (which is another case where the spatial weighting depends on the image data). Additionally, imaging parameters utilized during the capture of image data can also be considered in determining the weightings, such as (but not limited to) the gain or detected light level at which the image is captured, can be used to modulate the weighting factors.

The cost function $Cost_G$ (x,y,d) can be one of the metrics described above. In many embodiments, $Cost_G$ (x,y,d) uses a similarity measure based upon an L1 norm comparing a pixel in an alternate view image with a pixel in the reference image, an L2 norm comparing a pixel in an alternate view image with a pixel in the reference image, and/or variance across the pixels in the set of images captured by the Green cameras. In other embodiments, truncated versions of the L1 and L2 norms and/or any block-based similarity measure based on rank, census, correlation, and/or any other appropriate metric such as those practiced in multiview stereo disparity detection techniques can be utilized.

In a number of embodiments, the cost functions for the other color channels (i.e. $Cost_R$ (x,y,d) and $Cost_B$ (x,y,d)) do not utilize a comparison that includes a pixel from the reference image as the basis of determining pixel correspondence. In several embodiments, the similarity of corresponding pixels are performed by calculating the aggregated difference between each unique pair of corresponding pixels in the set of cameras within the color channel. In the example of an array camera in which depth is determined using four Red cameras, R_A, R_B, R_C, and R_D, the cost can be determined as follows:

$$Cost_R(x, y, d) = |R_A(x_A, y_A) - R_B(x_B, y_B)| + |R_A(x_A, y_A) - R_C(x_C, y_C)| + |R_A(x_A, y_A) - R_D(x_D, y_D)| + |R_B(x_B, y_B) - R_C(x_C, y_C)| + |R_B(x_B, y_B) - R_D(x_D, y_D)| + |R_C(x_C, y_C) - R_D(x_D, y_D)|$$

where (x_A, y_A), (x_B, y_B), (x_C, y_C), and (x_D, y_D) are pixel locations determined based upon the disparity in each of the cameras R_A, R_B, R_C, and R_D respectively at depth d.

The above metric can be referred to as the combination cost metric and can be applied within any color channel that does not contain the reference camera. In several embodiments, a combination metric can be utilized that does not include all combinations of unique pairs of corresponding pixels in the set of cameras within the color channel. In several embodiments, unique pairs of corresponding pixels from a subset of the images captured by an array camera can be utilized. When depths are determined for a virtual viewpoint, none of the spectral channels contain the "reference camera" and the combination cost metric can be applied in each of the spectral channels. Although the combination cost metric is shown above utilizing the L1 norm to determine the similarity between pixel intensity values, in other embodiments, the L2 norm, the pixel variance, truncated versions of the L1 and L2 norms and/or any block-based similarity measure based on rank, census, correlation, and/or any other appropriate metric such as those practiced in multiview stereo disparity detection techniques can be utilized.

Weighting factors (e.g. γ_G, γ_R, and γ_B) can be used to determine the contribution of each of the spectral channels to a depth estimate. The weights can be fixed or vary from pixel to pixel (i.e. spatially-varying) in the reference image. In many embodiments, a map of signal-to-noise ratio (SNR) can be generated with respect to the reference image using an SNR estimator. In several embodiments, the SNR estimator can determine SNR based upon a prior characterization of signal noise. Areas where the SNR response is high can indicate the presence of texture or high signal. Areas where the SNR estimate is low can indicate a textureless region, consisting almost entirely of noise. In certain situations, the data from the images might be noisy in certain spectral channels, but not in others. For example, an area may appear textureless in Green images, but have signal content in images captured by Red cameras. Therefore, the Green cameras will contribute little useful information to the cost function and may actually introduce noise into the depth estimation process, resulting in a less reliable depth estimate than if only Red or Blue cameras were included in the cost function. Therefore, an SNR map can be utilized to determine weightings to apply to each of the color channels. In several embodiments, if the SNR estimate in the reference image for a pixel (x, y) is low, meaning that the immediate region around pixel (x, y) is likely textureless and does not contain significant signal, then the weighting for the color channel containing the reference image should be reduced at the pixel (x, y).

In many embodiments, a stricter condition can also be used and/or used as an alternative in which the weighting for the spectral channel containing the reference image should be reduced at the pixel (x, y), when the SNR estimate in the reference image at a pixel (x, y) and for the radius of maximum parallax (along epipolar lines) in the reference image for all of the cameras show low SNR, then the weighting for the spectral channel containing the reference image should be reduced at the pixel (x, y). The radius of maximum parallax can be determined only with respect to pixels located along epipolar lines determined with respect to the other cameras in the camera array within the spectral channel. The stricter criterion acknowledges that though the SNR may be low at the pixel location (x, y) in the reference image, there may be content some distance away (less than a maximum parallax shift) from pixel (x, y) in another camera within the color channel containing the reference camera which could disqualify a candidate depth from being a likely match. Therefore, though the pixel location (x, y) may have low SNR, nearby content may still provide useful information to disqualifying certain depths as possibilities.

In a number of embodiments, strong SNR in the reference image may be used to reduce the weighting applied to the other color channels to save computation (i.e., fewer cameras must be searched). In addition, the SNR may be estimated for a camera in one of the other color channels to determine the weighting that should be applied to the color channel in the cost function. In many embodiments, the process of determining the SNR involves estimating SNR along the epipolar line which connects the pixel location (x, y) in the alternate view camera to the reference camera. Then, the epipolar line or line(s) may be searched for regions of high SNR. If high SNR contributions are found in the alternate view camera along the epipolar line, the weighting of the color channel to which the alternate view camera belongs can be set so that the color channel contributes to the cost metric for the pixel location (x, y) in the reference image. If, instead, along the epipolar line beginning at the pixel location (x, y), the alternate view image shows only low SNR, then the contribution of the color channel containing the alternate view image can be correspondingly reduced. In many embodiments, multiple cameras in each color channel are considered when determining the weighting to apply to the color channel when determining depth estimates. Although specific processes for esti-

mating depth using information contained within color channels that do not include the reference camera are described above, any of a variety of processes can be utilized to determine depth estimates based upon information contained in multiple color channels as appropriate to the requirements of specific applications in accordance with embodiments of the invention.

Based upon the initial depth map, visibility $V^{i,Ref}(x, y)$ can be updated based upon the computed depth map D(x,y) or based upon a filtered depth map F(D(x,y)) that is filtered using either a fixed coefficient filter (such as a Gaussian or a box filter), or adaptive or edge preserving filter such as (but not limited to) a joint bilateral filter that uses the reference image as guidance. A variety of techniques can be utilized for determining whether a pixel in an image captured from a reference viewpoint is occluded in another image. In a number of embodiments, an initial depth map is utilized to identify pixels that may be occluded. Information concerning foreground objects can be utilized to identify zones in which occlusions are likely to occur. In addition, any of a variety of additional characteristics of an initial depth map can be utilized to detect occlusions including (but not limited to) unusual depth transitions within the depth map and/or pixel depth values that are inconsistent with local pixel depth values.

Although much of the discussion above with respect to FIG. **8** relates to generation of depth estimates using image data in a single spectral channel or in the Red, Green, and Blue color channels, the techniques described above are equally appropriate with respect to any of a variety of spectral channels (the terms color channel and spectral channels being used herein interchangeably). An aggregated cost function similar to those described above can be utilized including a cost term determined with respect to each spectral channel using any of the techniques described above. In addition, the cost function can include cost terms that weight the similarity of pixels across spectral channels using techniques similar to the those described above. The accuracy of resulting depth estimates can depend upon the extent to which the depth estimate utilizes pixels from an image in which a pixel location (x, y) in a reference image is occluded. Techniques for improving the accuracy of depth estimates when a pixel location (x, y) in an image from a reference viewpoint are occluded within one or more images in the set of images are discussed further below.

Generating a Depth Map Accounting for Occlusions Using Subsets of Images

Patterns of visibility occur in natural scenes. Therefore, the pattern $V^{i,Ref}(x, y)$ is typically not arbitrary. A strong prior exists concerning the cameras in which a pixel in the reference image is visible. In many embodiments, a depth map can be determined in a manner that also provides an estimate for $V^{i,Ref}(x, y)$ in which there is a low likelihood that cameras in which pixel (x, y) is occluded are incorrectly identified. Based upon the estimate for $V^{i,Ref}(x, y)$, a depth estimate can be determined without the need to iterate to refine the visibility of $V^{i,Ref}(x, y)$. Alternatively, additional iterations can be performed to refine the depth map by including additional cameras based upon visibility information obtained using a reliable depth estimate. By obtaining a better initial depth map, however, the iterative process is likely to converge more rapidly.

In many embodiments, the process of generating a depth map involves determining depth using multiple clusters or subsets of cameras that each correspond to a different pattern of visibility within the scene and selecting the depth estimate as the depth determined using the subset of images captured by the cluster of cameras in which corresponding pixels have the highest similarity. A process for determining a depth for a pixel (x, y) using images captured by groups of cameras representing subsets of a camera array in accordance with an embodiment of the invention is illustrated in FIG. **8**A. The process **850** includes selecting (**852**) an initial group of cameras corresponding to a specific pattern of visibility within the scene and determining (**854**) a candidate depth estimate using the subset of image data captured by the group of cameras is generated based upon the depth at which corresponding pixels within the group of cameras have the highest similarity. In several embodiments, the process is similar to that outlined above with respect to FIG. **8** with the exception that the costs are determined for each subset of images and the lowest cost depth estimate generated using the subsets is selected as a candidate depth estimate for a relevant pixel location. As is discussed further below, the similarity of corresponding pixels within the subset of images at the candidate depth estimate is then compared against the similarity of corresponding pixels within other subsets of images at other candidate depth estimates to determine the candidate depth estimate that is most reliable within the context of a given application. In many embodiments, the candidate depth estimate that is selected as the depth estimate for the pixel location is determined based upon the subset of images having the most similar corresponding pixels at the candidate depth estimate.

In many embodiments, the group of cameras can include cameras from multiple color channels and a cost metric weighting similarity in each color channel is utilized to estimate the depth of pixel (x, y) in the reference image using the group of cameras. The process then iterates (**856**, **858**, **854**) through a plurality of different groups of cameras that each correspond to different patterns of visibility within the scene until a depth estimate is determined for each of the groups of cameras. The depth estimate for pixel location (x, y) in an image from the reference viewpoint can be obtained by selecting the depth estimate from the group of cameras in which the corresponding pixels at the estimated depth have the highest similarity. As noted above, similarity of pixels can be determined using a cost function that weights similarity of pixels in multiple spectral channels and/or across spectral channels. The subset of images in which the corresponding pixels at the estimated depth have the highest similarity corresponds to a specific pattern of visibility and provides an initial estimate of $V^{i,Ref}(x, y)$ that has a low likelihood of incorrectly identifying that the pixel location (x, y) in an image from the reference viewpoint is visible in a camera in which it is occluded.

Although the discussion provided above is presented in the context of performing searches with respect to each group of cameras with respect to pixel locations (x, y) in the reference image, depth maps can be separately estimated for the pixels in the reference image using each group of cameras corresponding to a specific visibility pattern. In this way, the cost metrics determined for pixel location (x, y) using a particular group of cameras can be filtered to smooth out noise in the cost function. Therefore, the depth estimate for the pixel location (x, y) can be selected using the depth estimate for the group of cameras having the smallest filtered cost metric. The filters applied to the cost metrics determined using a specific group of cameras can be fixed, or can be spatially adaptive. The specific filters utilized can be determined based upon the requirements of specific applications in accordance with embodiments of the invention. Following selection of the depth estimates for the pixels in the reference image, additional filtering can be performed to further smooth noise in the initial depth map.

The clusters or groupings of cameras utilized to detect particular patterns of visibility within a scene can depend upon the numbers of cameras in an array camera, the camera that is selected as the reference camera, and/or the distribution of cameras from different color channels within the array. Eight groups of cameras in a 5×5 array corresponding to different patterns of visibility that are likely to be present within a scene with respect to pixels in a reference camera located at the center of the array are shown in FIGS. **8**B-**8**I. The eight groups are generated by rotating and flipping the same group template, which includes 12 cameras. Depending upon the orientation of the group template, this includes seven Green cameras, and either three Red cameras and 2 Blue cameras, or 3 Blue cameras and 2 Red cameras. As noted above, the group template can be utilized to select groups of cameras when estimating depth for a pixel (x, y) in a reference Green camera located at the center of the 5×5 array. The depth of the pixel location (x, y) can be estimated by selecting the depth estimate from the group of cameras in which the corresponding pixels in the three color channels at the estimated depth have the highest similarity.

Although specific groups are shown in FIGS. **8**B-**8**I for selecting groups of cameras, any of a variety of templates corresponding to common visibility patterns within a scene can be utilized. For example, groups of cameras along a single epipolar line can be selected as described below with reference to FIG. **10**. In many embodiments the groups are selected so that the same number of cameras in the color channel containing the reference camera appears in each group of cameras. In addition, the groups can be determined so that there are at least two cameras in the other color channels in each group of cameras. If the groups include an uneven number of cameras, then the cost metric with respect to different sized groups may be biased and the bias can be accounted for through normalization. In many embodiments, the groups of cameras are selected to provide baseline diversity (contrast with the groups illustrated in FIG. **10** that are selected based upon sharing a common baseline). The greater the number of different radial epipolar lines on which depth searches are performed, the more likely one of the images captured by a group of cameras will contain information that can assist in identifying incorrect depths. In several embodiments, the group of cameras are selected so that the central angle of the sector defined by the epipolar lines of each group is the same.

In smaller array cameras, such as (but not limited to) 4×4 array cameras, and depending upon the pattern of color filters utilized within the array it may not be possible to select groups of cameras that contain the same number of cameras in each color channel. In several embodiments, a color filter pattern is utilized so that groups of cameras corresponding to common visibility patterns contain the same number of cameras in a single color channel. In this way, image data captured within the color channel can be utilized to estimate depths for occluded or otherwise mismatched pixels by comparing the filtered costs of depth estimates obtained using the different subgroups. Four groups of cameras in a 4×4 array corresponding to different patterns of visibility that are likely to be present within a scene with respect to pixels in a reference camera located at the center of the array are shown in FIGS. **8**J-**8**M. The four groups are generated by rotating and flipping the same group template, which includes 4 cameras. In the illustrated embodiment, there are three color channels: Red, Green, and Blue. Each group of cameras includes three Green cameras and one Blue or Red camera. Due to the presence of a single Red or Blue camera, in several embodiments depth estimates are determined using the image data captured in the Green color channel. For a given pixel location (x, y), the image data in the Red or Blue camera of the group that yielded the most reliable depth estimate (i.e. the lowest cost) is assumed visible in an image from the reference viewpoint. Accordingly, the pixel value in the pixel location in the Red or Blue image corresponding to the pixel location (x, y) in the image from the reference viewpoint can be utilized as a reference pixel for the purpose of calculating the visibility of corresponding pixels in other images within the Red or Blue color channels. For each of the groups shown in FIGS. **8**J-**8**M, one of the spectral channels is excluded from the group. The use of a π filter group can, however, be utilized to identify which of the images in the excluded color channel should be used as a reference image for the purpose of determining the visibility of pixels in the excluded color channel. When the viewpoint of a central camera of a π camera group is utilized as a reference viewpoint, two images in the excluded color channel will have been captured from viewpoints on opposite sides of the reference viewpoint. In typical natural scenes, a pixel location within an image from the reference viewpoint is likely to be visible in at least one of images captured by the adjacent cameras in the excluded color channel. In order to determine which (if any) of the images is most likely to contain a corresponding pixel to a pixel location in an image from the reference viewpoint that is visible, the similarity of the corresponding pixels within the two subgroups that contain the two images. Assuming that the corresponding pixels in at least one of the subgroups achieves a threshold level of similarity, then the image in the subgroup in which the corresponding pixels have the highest level of similarity can be selected as a reference image for the excluded color channel. In this way, the visibility of corresponding pixels in any image within the excluded color channel can be determined based upon its similarity with the corresponding pixel from the reference image for the excluded color channel. Where neither image captured from the adjacent viewpoints to the reference viewpoint reliably contain a visible pixel corresponding to a pixel location within an image from the reference viewpoint, then alternative techniques can be utilized to identify an image within the excluded color channel that contains a corresponding pixel that is visible and/or to determine the visibility of pixels within individual images from the excluded color channel. In several embodiments, visibility can be determined by performing epipolar line searches in the manner described herein. In a number of embodiments, cross-channel similarity measures can be used to determine a corresponding pixel within the images in an excluded color channel that can be utilized as a reference pixel. In several embodiments, the image in which the neighborhood surrounding the corresponding pixel exhibits the highest cross-correlation (or any other appropriate cross-channel similarity measure) with the reference image can be utilized as a reference pixel for the purpose of determining the visibility of the other corresponding pixels in the images in the excluded color channel. A similar approach can be taken with array cameras including different sized camera arrays.

In many embodiments, the groups of cameras used to estimate depth for a pixel in a reference camera correspond to pairs of cameras within the array. Through use of thresholds, cameras in which a pixel (x, y) is likely to be visible can be identified and an initial visibility map $V^{i,Ref}(x, y)$ constructed. The threshold can be a hard threshold, and/or a threshold based upon the SNR in the reference image. The same is also true of larger groups of cameras such as those illustrated in FIGS. **8**B-**8**I. Thresholds can be used to detect the presence of one or more outlier pixels within a set of corresponding pixels. Groups of cameras that are found to not contain out-

liers can then be combined, and the depth recalculated with this new combined set, to improve the precision of depth estimates. In a similar manner, an initial depth map can be constructed by initially assuming that all cameras are visible in the visibility map $V^{i,Ref}(x, y)$. Any of a variety of techniques can be utilized to determine whether a pixel (x, y) is occluded in at least one of the cameras in the camera array including (but not limited to) use of thresholds in the manner outlined above, and/or performing occlusion searches along epipolar lines. The depth of pixels that are likely to be occluded in at least one of the cameras in the array can then be estimated again using a process similar to the process outlined above with respect to FIG. **8**A. In this way, the cameras in which the pixel is occluded can be rapidly identified.

Although a variety of processes for determining depth maps and visibility maps when estimating depth for pixels within a reference image are described above with reference to FIGS. **8**A-**8**I, any of a variety of processes can be utilized to determine an initial depth map and/or visibility map in accordance with the requirements of specific applications in accordance with embodiments of the invention. Processes for identifying occluded pixels including processes that involve performing searches for occluded pixels along epipolar lines in accordance with embodiments of the invention are discussed further below.

Identifying Occluded Pixels

A challenge associated with identifying occluded pixels from an initial depth map is that depths within the depth map that are determined using occluded pixels may be incorrect. The most reliable depth estimates are those of the objects in the scene that are closest to the camera. These are the objects that also give rise to the greatest disparity and can potentially result in the largest number of pixel occlusions (depending upon the distribution of objects within the scene. Therefore, a determination can be made concerning whether a pixel visible in the reference image is occluded in a second image by searching for an occluding pixel in the reference image along the baseline vector. An occluding pixel is a pixel that is sufficiently close to the camera that the disparity observed from the perspective of the second image would be sufficiently large as to occlude the pixel visible in the reference image. The search for occluding pixels can be understood with reference to FIG. **9**. An image captured from a reference viewpoint **900** is shown in FIG. **9**. In order to determine the visibility of pixel (x_1, y_1) with a depth d_1 in a second image captured by the array camera, a search is conducted along a line **902** parallel to the baseline between the camera that captured the reference image and the camera that captured the second image. The pixel (x_1, y_1) will be occluded, when a pixel (x_2, y_2) is closer to the camera (i.e. located at a depth $d_2<d_1$). Therefore, all pixels (x_2, y_2) where $d_2 \geq d_1$ can be disregarded. Where the scene-dependent geometric shifts of each pixel (s_2 and s_1 respectively) are greater than the distance between the two pixels along the line **902** parallel to the baseline vector, then pixel (x_2, y_2) will also occlude the pixel (x_1, y_1). Stated another way, pixel (x_2, y_2) occludes pixel (x_1, y_1) when

$$|s_2 - s_1 - \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}| \leq \text{threshold}$$

In several embodiments, the threshold in the above expression can be determined as the inverse of the super-resolution factor used during subsequent super-resolution processing (e.g. when the super-resolution process attempts to achieve an increase in resolution of a factor of 3, then a threshold of ⅓ of a pixel is appropriate). When no pixel can be found satisfying the above expression, then the pixel (x_1, y_1) can be concluded to be visible in the second image. Where the above expression is satisfied for some pixel (x_2, y_2), then the pixel (x_1, y_1) can be considered to be occluded. Therefore, the process illustrated in FIG. **8** can be repeated to create an updated depth estimate for the pixel (x, y) disregarding the second image (and any other images in which the pixel is occluded). As can readily be appreciated, incorrect depths in the initial depth estimate can result in visible pixels being disregarded in future iterations of the process for determining a depth map. Using adaptive support to provide depths that are photometrically consistent decreases the likelihood that noise will result in the detection of false pixel occlusions, which eliminate useful information from subsequent process iterations. In many embodiments, the decision to designate a pixel as being occluded considers the similarity of the pixels and the confidence of the estimated depths of the pixels (x_1, y_1) and (x_2, y_2). As is discussed further below, a confidence map can be generated with respect to the depth map of the reference image and the confidence map indicates the reliability of a specific depth map. Therefore, a possible occlusion identified using the expression provided above in which scene-dependent geometric shifts of each pixel (s_2 and s_1 respectively) are based upon unreliable depth estimates can be disregarded. Similarly, a possible occlusion involving pixels where the difference in the intensities of the pixels is below a predetermined threshold can be disregarded. Where the pixels values are sufficiently similar, a depth estimate generated in reliance on the pixel will largely be unaffected even if the pixel is occluded. In other embodiments, a variety of other considerations can be taken into account when determining whether to indicate a pixel as occluded in a visibility map as appropriate to the requirements of specific applications.

The search discussed above with respect to FIG. **9** can be conducted along the epipolar line corresponding to every camera utilized to perform the depth estimation. When the captured images are rectified correctly, the search can be simplified by aligning the baselines of the cameras relative to the rows and columns of the pixels captured from the reference viewpoint (see discussion of rectification above). The search for occluding pixels need not be performed with respect to every pixel in the reference image. Instead, an initial search can be conducted for pixels that are likely to be occluded in one or more images captured by the array camera including (but not limited to) pixels proximate depth transitions. Searches can then be performed for occluding pixels with respect to pixels that are considered likely to be occluded. In addition, the search for occluding pixels can be performed more efficiently by computing the projections of pixels based upon distance in advance (the projections indicate the depth at which adjacent pixels along the baseline will be occluded). In addition, once a pixel is determined to be occluded the process for detecting occlusion of adjacent pixels can be simplified by utilizing the projection of the occluding pixel. In other embodiments, any of a variety of techniques can be utilized to more efficiently locate occluding and occluded pixels including (but not limited to) rendering the image based on depth in accordance with embodiments of the invention.

As noted above, including occluded pixels in the determination of the initial depth map can introduce errors in the resulting pixel depths. When occlusions are detected using a process similar to any of the processes outlined above and the depth map updated, errors in the depth map are removed. As errors in the depth map are removed, more accurate predictions can be made as to the pixels that are occluded. Accordingly, the process of detecting occlusions can be performed iteratively until a stopping criterion is reached. In a number of embodiments, the stopping criterion can be (but is not limited

to) the number of occluded pixels detected in a specific iteration (that were not previously detected) falling below a predetermined number and/or the completion of a predetermined number of iterations.

Referring back to FIG. **6**, a process for generating and refining a depth map in accordance with an embodiment of the invention is illustrated. In many instances, the processes for determining (**602**) the initial depth map will have a tendency to overestimate the disparity of occluded pixels. This has the effect of pushing occluded pixels into the foreground. Therefore, in a number of embodiments, a pixel that occludes another pixel can also be treated like an occluded pixel for the purposes of updating (**606**) the depth map. In this way, the depth of background pixels that have incorrect initial depth measurements can be detected and updated. As is discussed further below, the visibility of pixels that are ignored can be separately determined (**610**) once the depth map is finalized. In a number of embodiments, processes such as the process **600** illustrated in FIG. **6** also involve application of a bilateral filter to the depth map to help stabilize the depth map as the process iterates.

The accuracy of a depth estimate typically increases with the number of images in the light field captured by the array camera utilized in generating the depth estimate. Considering a smaller number of images can, however, reduce the computational complexity of obtaining a depth estimate. When a depth transition occurs, occlusions will typically occur in images captured on one side of the reference viewpoint. Therefore, a search for occluding pixels similar to the search described above can be utilized to determine whether a pixel is occluded in a group of images captured to one side of the reference camera. If the search indicates that no occlusions occurred, then the depth of the pixel can be determined using that group of images and the depth map updated.

A 5×5 array camera that can be utilized to construct a depth map using the Green cameras in the array is illustrated in FIG. **10**. The array camera **1010** includes a central reference Green camera (**1012**). The remaining Green cameras in the array can be utilized to form eight radial groups of three Green cameras for the purpose of determining depth of pixels that are occluded in at least one of the images captured by the Green cameras. Although radial groups are illustrated in FIG. **10**, groups of cameras in each of four quadrants surrounding the reference viewpoint can also be utilized. A group may be as small as a pair of cameras, one of which is the camera that captures an image from the reference viewpoint. In many embodiments, groups such as those discussed above with reference to FIGS. **8**A-**8**I can also be utilized.

Although specific processes for detecting pixel occlusions are discussed above with respect to FIGS. **6**, **8**A-**8**I, **9**, and **10**, any of a variety of processes can be utilized to generate a depth map including (but not limited to) processes that reduce the computational complexity of detecting occlusions in accordance with embodiments of the invention. In several embodiments, the process of determining the depth of each pixel can involve searching based upon both hypothesized depths and hypothesized visibility and the combination of depth and visibility that yields the highest pixel correspondence selected as the most likely depth and set of occlusions. Visibility determined in this way can be confirmed by using the approach described above for detecting occluding pixels.

In many embodiments, information concerning the visibility of pixels in the captured images from the reference viewpoint is utilized in processes including (but not limited to) super-resolution processing. Processes for determining the visibility of pixels in images captured by an array camera from a reference viewpoint using a depth map in accordance with embodiments of the invention are discussed further below.

Determining Visibility of Pixels

Pixel visibility can be utilized in determining a depth map and in a variety of other applications including (but not limited to) super-resolution processing. In several embodiments, a depth map for an image captured from a reference viewpoint generated using a process similar to the processes outlined above is utilized to generate a visibility map for the other images captured by an array camera (i.e. the images captured from alternate viewpoints). In several embodiments, visibility maps can be determined with respect to whether pixels in alternate view images are visible from the reference viewpoint and whether a pixel in a first alternate view image is visible any of the other alternate view images. In a number of embodiments, the process of determining visibility maps for the images captured within a single color channel involves comparing the photometric similarity of pixels corresponding to a pixel in the image captured from the reference viewpoint. Pixels that are considered to have a predetermined level of similarity can be considered visible and pixels that are below a threshold level of similarity can be considered occluded. The threshold utilized to determine the photometric similarity of corresponding pixels can adapt based upon the similarity of the corresponding pixels. In several embodiments, the threshold is determined as a function of the photometric distance between the pixel from the reference image and the corresponding pixel that is most similar to the pixel from the reference image. When an array captures image data in multiple color channels, the visibility of pixels in a single color channel can be utilized to determine the visibility of pixels in other channels.

A process for determining the visibility of corresponding pixels to a pixel within a reference image in accordance with an embodiment of the invention is illustrated in FIG. **11**. The process **1100** includes selecting (**1102**) a pixel from an image captured from the reference viewpoint. A depth map generated using a process similar to the processes described above can be utilized to determine the depth of the selected pixel. Based upon the depth of the selected pixel, the locations of the corresponding pixels in each image captured by the array camera can be determined (**1104**). The similarity of the corresponding pixels to the selected pixel from the reference image can be utilized to determine the visibility of the corresponding pixels. In a number of embodiments, the photometric distance of the pixels is utilized as a measure of similarity and a threshold used to determine pixels that are likely visible and pixels that are likely occluded. In many embodiments, the threshold varies depending upon the characteristics of the pixels that are compared. In certain embodiments, the threshold value used to determine similarity of corresponding pixels is determined using the intensity of a reference pixel, as the average of a subset of pixel intensity values for corresponding pixels that are found to be visible in a given color channel. In several embodiments, the specific corresponding pixel intensities that are averaged can depend upon corresponding camera baseline and confidence values for the pixels (if available) and associated matching costs for the pixels. In several embodiments, the threshold is determined (**1106**) as a function of the photometric distance between the selected pixel from the reference image and the corresponding pixel that is photometrically closest to the selected pixel. In a number of embodiments, the threshold is based upon the pixel intensity of the corresponding pixel in the reference image and/or the intensity of the pixel in the alternate view image. In certain embodiments, the threshold is determined using an SNR

model for the captured image. In a number of embodiments, the photometric distance of the selected pixel and the closest corresponding pixel is scaled and/or an offset is added to obtain an appropriate threshold. In other embodiments, any of a variety of techniques can be utilized for determining a threshold for determining the visibility of a corresponding pixel including (but not limited to) using a fixed threshold.

The selected pixel from the reference image and the corresponding pixels are compared (**1108**) and the threshold used to determine (**1110**) the similarity of the pixels. When the photometric distance of the selected pixel from the reference image and one of the corresponding pixels is less than the threshold, then the corresponding pixel is determined (**1112**) to be visible. When the photometric distance of the selected pixel from the reference image and one of the corresponding pixels exceeds the threshold, then the corresponding pixel is determined (**1114**) to be occluded.

The process (**1100**) illustrated in FIG. **11** can be repeated for a subset or all of the pixels in the reference image to generate visibility maps for the corresponding pixels in other images captured by the array camera. In embodiments where all of the pixels in the camera that captures an image from the reference viewpoint are in a single color channel, then processes similar to the process illustrated in FIG. **11** effectively generate visibility for images captured within a single color channel. When the array camera also includes images captured in other color channels, the visibility of pixels in images that are not in the same color channel as the reference image can be determined by performing similar comparisons to those described above with respect to corresponding pixels from images within the spectral channel that known or are likely visible in the reference viewpoint. In other embodiments, the camera that captures the reference image employs a Bayer filter (or another appropriate filter pattern) that enables the capture of image data in multiple color channels from the reference viewpoint. In which case, processes similar to those illustrated in FIG. **11** can be utilized to generate visibility information for images in multiple color channels, where the process involves demosaicing the Bayer filter pattern to obtain a Red and Blue pixel value at each position in the reference view.

Although specific processes are discussed above in the context of FIG. **11**, any of a variety of processes can be utilized to determine the visibility of pixels in images captured by an array camera in accordance with embodiments of the invention including (but not limited to) processes that iteratively refine visibility information as part of the process of generating a depth map. In many embodiments, the process of generating a depth map and a visibility map also includes generating a confidence map that can provide information concerning the reliability of the estimated depths within the confidence map. Processes for determining confidence maps in accordance with embodiments of the invention are discussed further below.

Confidence Maps

Processes for generating depth maps, including those described above, can result in regions within a depth map in which depth estimates are unreliable. A textureless region of an image synthesized using image data captured by an array camera is illustrated in FIG. **18**A and the depth map for the image generated using processes similar to those described above in accordance with embodiments of the invention is illustrated in FIG. **18**B. In the textureless region **1800**, the cost metric used to determine depth in the manner described above is noisy and though a minimum cost (i.e., at a depth where the cameras show maximum correspondence) can be found, the result depends largely on sensor and photon noise and not any significant underlying signal. The pixel correspondence in such regions (as measured by the cost metric) is indeed greatest at the depth shown, but the resulting depth shown is not the correct depth of the object. In contrast, in the edge region **1802**, the cost function shows a depth at which the cost is minimized with great certainty. There, the edge signal is much greater than the noise and so the disparity corresponding to the actual depth of the object can be detected with higher confidence.

Depth errors are not limited to textureless regions, however. Another class of depth errors occurs in zones of occlusion, where certain background regions are visible in some cameras, and not others. This sort of error can be seen along the rim of the tire, where the foreground region intersects the background region. In the depth map, there appear to be regions **1804** containing incorrect depth estimates at the interface between the foreground to background.

When generating a depth map, a confidence map can be generated, which describes numerically, through some measure, the reliability of different depth estimates within the depth map. The confidence map can be used by later processing stages, or by third-party applications, to determine which regions of the depth map can be most relied upon for further processing. For example, a depth measurement utility can allow a user to click on regions of an image synthesized using a super-resolution process to obtain the depth of a particular pixel. If the user clicks on a pixel of the image, the confidence map can be checked before returning a result. If the confidence of the depth at the requested location is low, then the user interface can avoid reporting the depth at that location. If the confidence map is high, then the user interface can safely report the depth at the selected location. That is, the confidence map can be used to qualify the results for particular applications and not return an inaccurate value where the depth map is known to contain errors.

The confidence map can be encoded in a variety of ways, and there may be multiple classes or axes of confidence encoded within a single confidence map. A variety of confidence measures that can be utilized to encode a confidence metric in a confidence map in accordance with embodiments of the invention are discussed below.

In several embodiments, a confidence map is encoded with a confidence measure based on whether the depth estimate of a pixel is within a textureless region within an image. As noted above, textureless regions can be detected based upon SNR in the region surrounding a given pixel. If the SNR is above a known tunable threshold, the region may be marked textureless in a binary fashion. Alternatively, the SNR itself (without being thresholded) may be remapped linearly or non-linearly and used to serve as a continuous indicator of confidence.

In many embodiments, a gradient edge map (e.g. Prewitt or Canny) may be calculated and used as a confidence metric within a confidence map. Since edges and texture typically have high confidence, gradient maps and SNR maps often provide a good coarse measure of confidence for a depth map.

In a number of embodiments, the confidence map can be encoded based upon whether a particular region is low confidence due to occlusions and/or mismatch and conflicting measurements between cameras (i.e. there may be texture in a region that is detected by an SNR map, but there may still be a depth error occurring because in that area the parallax detection process detects and/or is unable to resolve occlusions or otherwise confidently estimate depth for any other reason).

In a number of embodiments, a confidence metric is determined as the "best cost" achieved during the depth estimation

process, or a linear, non-linear, or quantized remapping of this quantity. If the minimum cost achieved during depth estimation is above a selected threshold, the region may be marked low confidence on the basis of a lack of correspondence between multiple views at the estimated depth.

In a number of embodiments, occlusions may be detected by comparing the best costs between different subgroups or depth maps generated between different groups of cameras and if the difference between best costs is greater than a threshold, marking low confidence for the pixel locations where the costs found in subsets of images differ.

In a number of embodiments, the confidence map can be encoded based upon whether a particular region is low confidence due to adaptive processing steps in the depth estimation process itself. For example, if fewer depths were searched in a particular region, this information can be encoded numerically in the confidence map to highlight that the depth is less reliable. In many embodiments, certain regions of the depth map are actually searched through correspondence search, and other regions of the depth map, the depths are interpolated based on results from a depth search on a sparser set of points. In such a case, the pixels with interpolated depths are given lower confidence values than pixels where the correspondences have actually been searched.

In several embodiments, the expected precision of a depth measurement can also be coded in the confidence map as a numerical quantity. In many instances, depths farther away from the camera are measured with greater error and so should be less trusted. In such cases the confidence map can mark such areas as involving lower confidence depth estimates. The confidence can be proportional to the expected depth error between adjacent search positions at that depth. In certain embodiments, the disparity corresponding to the minimum distance supported by the parallax search (i.e. this will be the maximum disparity observed between any two cameras for all supported depths) can be determined. Once the maximum disparity is found, the search will search a number of disparities up to the maximum disparity. In many embodiments the maximum disparity is D low resolution pixels and the number of depths searched is N. The number of pixels of disparity between adjacent search positions along an epipolar line is D/(N−1). The depth in meters for any one of the N disparities that is searched (indexed by n<N) are d_n=C/(n*D/(N−1)) where C is a constant that incorporates information about the baselines and focal lengths of the individual low resolution cameras having the maximum baselines. If, at a particular point in the depth map, the depth is d_n, then the expected measurement error at d_n is e_n=½*max($|d_n-d_{n+1}|$, $|d_n-d_{n-1}|$). Basically, the expected measurement error is the error due to searching a fixed, discrete number of points along the epipolar line. The error value itself may be mapped linearly or non-linearly to provide a confidence value with respect to a depth estimate at a particular pixel location within the depth map. The higher the error, the less confident the depth estimate. In several embodiments, disparities searched are not spaced equally, but may be coarser in some regions than others. Accordingly, error can be calculated similarly between adjacent indices (whatever the distribution of search positions along the epipolar line) so that the confidence map reflects the calculated error in depth. In a number of embodiments, the confidence map reflects the maximum error in estimated disparity (not depth), the inverse of the quantity listed above. This may be more useful for applications such as image fusion, whereas the depth error would be more useful for measurement applications that occur in real world coordinates (such as, but not limited to, 3D modeling).

In a number of embodiments, the confidence map can mark regions as low confidence due to known or detected lens or sensor defects that make the depth map unreliable. Defective pixels (a term that includes both defective pixels on the sensor as well as pixels affected by lens defects) may be detected during image processing or offline in a pre-processing calibration step. In one embodiment, if the total number of pixel defects within a radius of a particular pixel (x, y) in any reference camera exceeds a pre-set threshold, the pixel (x, y) is marked low confidence in the depth map due to pixel defects. In another embodiment, a similar confidence value may be defined where confidence increases proportionally (not as a hard threshold) as a function of the number of pixel defects in any camera within a radius and/or region surrounding the reference pixel (x, y) (or pixels known to be affected by lens defects). In another embodiment, the confidence may be a pre-calculated value for specific configurations of defective pixels that are known to create errors of varying severity. In several embodiments the confidence value for the defect takes into account the local image content in calculating the confidence value.

In several embodiments, the confidence map may mark as low confidence areas where the reference image appears textureless but in other color channels there is textured content. In one embodiment, a pixel (x, y) in the reference camera is searched within a local radius and/or region. If within this local radius and/or region the content is considered to be textureless in Green, but if the same search within another (perhaps larger) local radius/region for Red and/or Blue cameras turns up sufficient texture in images within the Red and/or Blue color channels, the region can be marked as lower confidence due to the fact that the Green color channel is less useful in this detection scenario and depth results will be less reliable (though often correct).

In a number of embodiments the confidence map numerically encodes as low confidence scenarios in which there is photometric mismatch due to lens flare or features in the scene. In many embodiments, the local statistics (mean and variance) of a region of interest around the pixel location (x, y) may be calculated and compared to the local statistics of a similar region in another camera. In this way, local image statistics between two neighborhoods in the same general region of multiple images can be compared to detect possible lens flare, the presence of which reduces confidence. In other embodiments, any of a variety of techniques can be utilized to compare neighborhoods in multiple images to detect lens flare. The resulting confidence measure can be a scaled or non-linearly mapped function of the difference between the mean and variance of the regions across images captured by multiple cameras. The greater the mean and variance differences between the images captured by the cameras, the less likely the depth is reliable and the lower the confidence will be.

In a number of embodiments the confidence map adapts to lighting conditions to reduce the confidence when the image is very dark and noisy. In certain embodiments, the sensor gain at the time the image was taken will result in an absolute reduction in confidence for all depths. In another embodiment, the analog gain and exposure time of the sensor are taken into account when computing a SNR ratio, or thresholds for edge gradients at different levels of noise. In many embodiments, the analog gains and exposure times for different focal planes can be utilized for different cameras in a camera array used to capture a set of images.

To detect regions which are of low confidence due to occlusions, the best-achieved cost metric may be stored during the parallax search. For regions which show significant occlu-

sion, the best achieved cost metric usually greatly exceeds the minimum value that would occur if there were no occlusion and all cameras saw the same content. Accordingly, a threshold can be applied to the best achieved costs. If the best achieved cost exceeds the threshold, then the region is marked as likely to have been occluded or to have had some other problem (such as photometric non-uniformity).

For certain similarity metrics, the low-confidence threshold for occlusion can be corrected for the mean intensity of the region as well as the noise statistics of the sensor. In many embodiments, the mean of the region in the reference image is calculated using a spatial box N×N averaging filter centered around the pixel of interest. In other embodiments, once the mean is known, the noise statistics for the color channel containing the reference camera may be calculated by a table lookup which relates a particular mean at a particular exposure and gain to a desired threshold. If the best matching value greatly exceeds the expected noise, then the pixel can be marked as unreliable due to possible occlusion.

A non-binary measure of confidence due to general mismatch can be obtained using the following formula:

$$\text{Confidence}(x,y)=F(\text{Cost}_{min}(x,y),\text{Cost}^{d}(x,y),I(x,y)^{cam},\text{Sensor},\text{Camera intrinsics})$$

where $\text{Cost}_{min}(x, y)$ is the minimum cost of a disparity search over the desired depth range,

$\text{Cost}^{d}(x, y)$ denotes that cost data from any depth or depths (beside the minimum depth),

$I(x, y)^{cam}$ image data captured by any camera can be utilized to augment the confidence;

Sensor is the sensor prior, which can include known properties of the sensor, such as (but not limited to) noise statistics or characterization, defective pixels, properties of the sensor affecting any captured images (such as gain or exposure),

Camera intrinsics is the camera intrinsic, which specifies elements intrinsic to the camera and camera array that can impact confidence including (but not limited to) the baseline separation between cameras in the array (affects precision of depth measurements), and the arrangement of the color filters (affects performance in the occlusion zones in certain scenarios).

In several embodiments, Confidence(x, y) may make use of values neighboring pixel location (x, y) (i.e. spatial neighborhoods) for all the arguments.

In many embodiments, a confidence map can be encoded based upon factors including (but not limited to) one or more of the above factors. Each factor may be encoded in a binary fashion or may be represented as a range of (quantized) degrees of confidence, or may be non-quantized ranges or derivatives thereof. For example, the confidence along a particular axis may be represented not as a single bit, but as multiple bits which represent the level of confidence that a region is textureless. In certain embodiments the confidence along a particular axis may be represented as a continuous or approximately continuous (i.e. floating point) quantity. Other factors considered when determining confidence can be determined using any of a variety of ranges of degrees of confidence as appropriate to the requirements of specific applications in accordance with embodiments of the invention. In several embodiments, an arbitrary number of confidence codes or values are included in a confidence map for a particular pixel where one may specify any or all of these conditions. Specific confidence metrics are discussed further below.

In a particular embodiment where only the minimum cost is considered and noise statistics of the sensor follow a linear model, a simplified form may be used:

$$\text{Confidence}(x, y) = a \times \frac{Cost_{\min}(x, y)}{Avg(x, y)} + \text{offset}$$

where Avg(x,y) is the mean intensity of the reference image in a spatial neighborhood surrounding (x, y), or an estimate of the mean intensity in the neighborhood, that is used to adjust the confidence based upon the intensity of the reference image in the region of (x, y),

a and offset are empirically chosen scale and offset factors used to adjust the confidence with prior information about the gain and noise statistics of the sensor.

In this case, higher values would indicate lower confidence, and it would be up to the image processing pipeline to determine how to threshold the results.

In general, the confidence map provides metadata describing the depth estimates contained within the depth map that quantifies numerically the accuracy of detected depths in the image. In many embodiments, the confidence map may be provided in an external delivery format along with the depth map for use with the depth map in applications including (but not limited to) machine vision, gesture recognition, post capture image refocusing, real-time applications, image special effects, super-resolution, or other applications. An example of the manner in which a confidence map can be utilized in a depth estimation process to filter a depth map in accordance with an embodiment of the invention is illustrated in FIGS. **18**C-**18**H. Turning first to FIG. **18**C is an image of a scene containing objects at different depths synthesized using a super-resolution process from multiple images captured in different color channels (specifically Red, Green and Blue color channels). A depth map generated from the reference viewpoint using processes similar to those outlined above is illustrated in FIG. **18**D. As can be readily appreciated, the depth map is noisy. A confidence map generated using any of a variety of the metrics outlined above can be generated as part of the process of generating a depth map. A binary confidence map generated by thresholding SNR in accordance with an embodiment of the invention is illustrated in FIG. **18**E. An 8-bit confidence map generated based upon SNR in accordance with an embodiment of the invention is illustrated in FIG. **18**F. A binary confidence map generated by combining a confidence factor generated by thresholding SNR and a confidence factor generated by thresholding the number of corresponding pixels that are occluded in accordance with an embodiment of the invention is illustrated in FIG. **18**G. In several embodiments, the confidence map can be utilized to filter the depth map. A depth map that is filtered based upon a binary confidence map generated by thresholding SNR in accordance with an embodiment of the invention is illustrated in FIG. **18**H. Comparing the depth maps shown in FIGS. **18**D and **18**E reveals the value of the use of the confidence map in interpreting depth information generated using any depth estimation process. Although a specific confidence metric and filtering process are described above with reference to FIGS. **18**C-**18**H, any of a variety of confidence metrics can be utilized in the filtering and/or additional processing of depth estimates and/or depth maps in accordance with embodiments of the invention. The generation of confidence maps and the use of confidence maps to filter depth maps in accordance with embodiments of the invention is further illustrated in the close up images shown in FIGS.

18I-18L. With specific regard to FIG. 18I, a close up image of an object synthesized from a light field of images captured in Red, Green, and Blue color channels (each image containing image data in a single color channel) using super-resolution processing is shown. A depth map generated using the techniques outlined above is illustrated in FIG. 18J. A binary confidence map generated by thresholding SNR generated in accordance with an embodiment of the invention is illustrated in FIG. 18K. A multibit resolution confidence map generated in accordance with an embodiment of the invention using SNR is illustrated in FIG. 18L. A binary confidence map generated by thresholding the SNR of the region surrounding each pixel and by thresholding the number of pixels in the images within the light field that correspond to a pixel location in the image from the reference viewpoint that are occluded in accordance with an embodiment of the invention is illustrated in FIG. 18M. A depth map filtered using the binary confidence map shown in FIG. 18M is illustrated in FIG. 18N.

In several embodiments, a confidence map generated using one or more of the metrics described above can be inserted as an additional channel of information into the JPEG-DZ file format or other file formats. In several embodiments, the confidence map is encoded and decoded using processes similar to those outlined in U.S. patent application Ser. No. 13/631,731 to Venkataraman et al. entitled "Systems and Methods for Encoding Light Field Image Files", filed Sep. 28, 2012. The disclosure of U.S. patent application Ser. No. 13/631,731 is herein incorporated by reference in its entirety. Although specific metrics for determining confidence are described above, any of a variety of metrics for determining confidence appropriate to the requirements of a specific application can be utilized in accordance with embodiments of the invention.

Encoding and Decoding of Confidence Maps

A variety of container file formats including the JPEG File Interchange Format (JFIF) specified in ISO/IEC 10918-5 and the Exchangeable Image File Format (Exif) and can be used to store a JPEG bitstream. JFIF can be considered a minimal file format that enables JPEG bitstreams to be exchanged between a wide variety of platforms and applications. The color space used in JFIF files is YCbCr as defined by CCIR Recommendation 601, involving 256 levels. The Y, Cb, and Cr components of the image file are converted from R, G, and B, but are normalized so as to occupy the full 256 levels of an 8-bit binary encoding. YCbCr is one of the compression formats used by JPEG. Another popular option is to perform compression directly on the R, G and B color planes. Direct RGB color plane compression is also popular when lossless compression is being applied.

A JPEG bitstream stores 16-bit word values in big-endian format. JPEG data in general is stored as a stream of blocks, and each block is identified by a marker value. The first two bytes of every JPEG bitstream are the Start Of Image (SOI) marker values FFh D8h. In a JFIF-compliant file there is a JFIF APP0 (Application) marker, immediately following the SOI, which consists of the marker code values FFh E0h and the characters JFIF in the marker data, as described in the next section. In addition to the JFIF marker segment, there may be one or more optional JFIF extension marker segments, followed by the actual image data.

Overall the JFIF format supports sixteen "Application markers" to store metadata. Using application markers makes it is possible for a decoder to parse a JFIF file and decode only required segments of image data. Application markers are limited to 64K bytes each but it is possible to use the same maker ID multiple times and refer to different memory segments.

An APP0 marker after the SOI marker is used to identify a JFIF file. Additional APP0 marker segments can optionally be used to specify JFIF extensions. When a decoder does not support decoding a specific JFIF application marker, the decoder can skip the segment and continue decoding.

One of the most popular file formats used by digital cameras is Exif. When Exif is employed with JPEG bitstreams, an APP1 Application marker is used to store the Exif data. The Exif tag structure is borrowed from the Tagged Image File Format (TIFF) maintained by Adobe Systems Incorporated of San Jose, Calif.

In many embodiments, a light field image file is created by encoding an image synthesized from light field image data and combining the encoded image with a depth map derived from the light field image data. In several embodiments, the encoded image is synthesized from a reference viewpoint and the metadata includes information concerning pixels in the light field image that are occluded from the reference viewpoint. In a number of embodiments, the metadata can also include additional information including (but not limited to) auxiliary maps such as confidence maps, edge maps, and missing pixel maps that can be utilized during post processing of the encoded image to improve the quality of an image rendered using the light field image data file.

In many embodiments, the light field image file is compatible with the JPEG File Interchange Format (JFIF). The synthesized image is encoded as a JPEG bitstream and stored within the file. The accompanying depth map, occluded pixels and/or any appropriate additional information including (but not limited to) auxiliary maps are then stored within the JFIF file as metadata using an Application marker to identify the metadata. A legacy rendering device can simply display the synthesized image by decoding the JPEG bitstream. Rendering devices in accordance with embodiments of the invention can perform additional post-processing on the decoded JPEG bitstream using the depth map and/or any available auxiliary maps. In many embodiments, the maps included in the metadata can also be compressed using lossless JPEG encoding and decoded using a JPEG decoder. Although much of the discussion that follows references the JFIF and JPEG standards, these standards are simply discussed as examples and it should be appreciated that similar techniques can be utilized to embed metadata derived from light field image data used to synthesize the encoded image within a variety of standard file formats, where the synthesized image and/or maps are encoded using any of a variety of standards based image encoding processes.

By transmitting a light field image file including an encoded image, and metadata describing the encoded image, a rendering device (i.e. a device configured to generate an image rendered using the information within the light field image file) can render new images using the information within the file without the need to perform super resolution processing on the original light field image data. In this way, the amount of data transmitted to the rendering device and the computational complexity of rendering an image is reduced. In several embodiments, rendering devices are configured to perform processes including (but not limited to) refocusing the encoded image based upon a focal plane specified by the user, synthesizing an image from a different viewpoint, and generating a stereo pair of images.

Capturing and Storing Light Field Image Data

Processes for capturing and storing light field image data in accordance with many embodiments of the invention involve

capturing light field image data, generating a depth map from a reference viewpoint, and using the light field image data and the depth map to synthesize an image from the reference viewpoint. The synthesized image can then be compressed for storage. The depth map and additional data that can be utilized in the post processing can also be encoded as metadata that can be stored in the same container file with the encoded image.

A process for capturing and storing light field image data in accordance with an embodiment of the invention is illustrated in FIG. **25**. The process **2500** includes capturing (**2502**) light field image data. In several embodiments, the light field image data is captured using an array camera similar to the array cameras described above. In other embodiments, any of a variety of image capture device(s) can be utilized to capture light field image data. The light field image data is used to generate (**2504**) a depth map.

The light field image data and the depth map can be utilized to synthesize (**2506**) an image from a specific viewpoint. In many embodiments, the light field image data includes a number of low resolution images that are used to synthesize a higher resolution image using a super resolution process.

In order to be able to perform post processing to modify the synthesized image without the original light field image data, metadata can be generated (**2508**) from the light field image data, the synthesized image, and/or the depth map. The metadata data can be included in a light field image file and utilized during post processing of the synthesized image to perform processing including (but not limited to) refocusing the encoded image based upon a focal plane specified by the user, and synthesizing one or more images from a different viewpoint. In a number of embodiments, the auxiliary data includes (but is not limited to) pixels in the light field image data occluded from the reference viewpoint used to synthesize the image from the light field image data, one or more auxiliary maps including (but not limited to) a confidence map, an edge map, and/or a missing pixel map. Auxiliary data that is formatted as maps or layers provide information corresponding to pixel locations within the synthesized image. A confidence map is produced during the generation of a depth map and reflects the reliability of the depth value for a particular pixel. This information may be used to apply different filters in areas of the image and improve image quality of the rendered image. An edge map defines which pixels are edge pixels, which enables application of filters that refine edges (e.g. post sharpening). A missing pixel map represents pixels computed by interpolation of neighboring pixels and enables selection of post-processing filters to improve image quality. As can be readily appreciated, the specific metadata generated depends upon the post processing supported by the image data file. In a number of embodiments, no auxiliary data is included in the image data file.

In order to generate an image data file, the synthesized image is encoded (**2510**). The encoding typically involves compressing the synthesized image and can involve lossless or lossy compression of the synthesized image. In many embodiments, the depth map and any auxiliary data are written (**2512**) to a file with the encoded image as metadata to generate a light field image data file. In a number of embodiments, the depth map and/or the auxiliary maps are encoded. In many embodiments, the encoding involves lossless compression.

Although specific processes for encoding light field image data for storage in a light field image file are discussed above, any of a variety of techniques can be utilized to process light field image data and store the results in an image file including but not limited to processes that encode low resolution images captured by an array camera and calibration information concerning the array camera that can be utilized in super resolution processing. Storage of light field image data in JFIF files in accordance with embodiments of the invention is discussed further below.

Image Data Formats

In several embodiments, the encoding of a synthesized image and the container file format utilized to create the light field image file are based upon standards including but not limited to the JPEG standard (ISO/IEC 10918-1) for encoding a still image as a bitstream and the JFIF standard (ISO/IEC 10918-5). By utilizing these standards, the synthesized image can be rendered by any rendering device configured to support rendering of JPEG images contained within JFIF files. In many embodiments, additional data concerning the synthesized image such as (but not limited to) a depth map and auxiliary data that can be utilized in the post processing of the synthesized image can be stored as metadata associated with an Application marker within the JFIF file. Conventional rendering devices can simply skip Application markers containing this metadata. Rendering device in accordance with many embodiments of the invention can decode the metadata and utilize the metadata in any of a variety of post processing processes.

A process for encoding an image synthesized using light field image data in accordance with the JPEG specification and for including the encoded image and metadata that can be utilized in the post processing of the image in a JFIF file in accordance with an embodiment of the invention is illustrated in FIG. **26**. The process **2600** includes encoding (**2602**) an image synthesized from light field image data in accordance with the JPEG standard. The image data is written (**2604**) to a JFIF file. A depth map for the synthesized image is compressed (**2606**) and the compressed depth map and any additional auxiliary data are written (**2608**) as metadata to an Application marker segment of the JFIF file containing the encoded image. Where the auxiliary data includes maps, the maps can also be compressed by encoding the maps in accordance with the JPEG standard. At which point, the JFIF file contains an encoded image and metadata that can be utilized to perform post processing on the encoded image in ways that utilize the additional information captured in the light field image data utilized to synthesize the high resolution image (without the need to perform super resolution processing on the underlying light field image data).

Although specific processes are discussed above for storing light field image data in JFIF files, any of a variety of processes can be utilized to encode synthesized images and additional metadata derived from the light field image data used to synthesize the encoded images in a JFIF file as appropriate to the requirements of a specific application in accordance with embodiments of the invention. The encoding of synthesized images and metadata for insertion into JFIF files in accordance with embodiments of the invention are discussed further below. Although much of the discussion that follows relates to JFIF files, synthesized images and metadata can be encoded for inclusion in a light field image file using any of a variety of proprietary or standards based encoding techniques and/or utilizing any of a variety of proprietary or standards based file formats.

Encoding Images Synthesized from Light Field Image Data

An image synthesized from light field image data using super resolution processing can be encoded in accordance with the JPEG standard for inclusion in a light field image file in accordance with embodiments of the invention. The JPEG standard is a lossy compression standard. However, the information losses typically do not impact edges of objects. There-

fore, the loss of information during the encoding of the image typically does not impact the accuracy of maps generated based upon the synthesized image (as opposed to the encoded synthesized image). The pixels within images contained within files that comply with the JFIF standard are typically encoded as YCbCr values. Many array cameras synthesize images, where each pixel is expressed in terms of a Red, Green and Blue intensity value. In several embodiments, the process of encoding the synthesized image involves mapping the pixels of the image from the RGB domain to the YCbCr domain prior to encoding. In other embodiments, mechanisms are used within the file to encode the image in the RGB domain. Typically, encoding in the YCbCr domain provides better compression ratios and encoding in the RGB domain provides higher decoded image quality.

Storing Additional Metadata Derived from Light Field Image Data

The JFIF standard does not specify a format for storing depth maps or auxiliary data generated by an array camera. The JFIF standard does, however, provide sixteen Application markers that can be utilized to store metadata concerning the encoded image contained within the file. In a number of embodiments, one or more of the Application markers of a JFIF file is utilized to store an encoded depth map and/or one or more auxiliary maps that can be utilized in the post processing of the encoded image contained within the file.

A JFIF Application marker segment that can be utilized to store a depth map, individual camera occlusion data and auxiliary map data in accordance with an embodiment of the invention is illustrated in FIG. **27**. The APPS Application marker segment **2700** uses a format identifier **2702** that uniquely identifies that the Application marker segment contains metadata describing an image synthesized using light field image data. In a number of embodiments, the identifier is referred to as the "DZ Format Identifier" **2702** and is expressed as the zero terminated string "PIDZ0".

The Application marker segment includes a header **2704** indicated as "DZ Header" that provides a description of the metadata contained within the Application marker segment. In the illustrated embodiment, the "DZ Header" **2704** includes a DZ Endian field that indicates whether the data in the "DZ Header" is big endian or little endian. The "DZ Header" **2704** also includes a "DZ Selection Descriptor".

An embodiment of a "DZ Selection Descriptor" is illustrated in FIG. **28**, which includes four bytes. The first two bytes (i.e. bytes 0 and 1) contain information concerning the metadata describing the encoded image that are present (see FIG. **29**) and the manner in which the different pieces of metadata are compressed (see FIG. **30**). In the illustrated embodiment, the types of metadata that are supported are a depth map, occluded pixel data, virtual view point data, a missing pixel map, a regular edge map, a silhouette edge map, and/or a confidence map. In other embodiments, any of a variety of metadata describing an encoded image obtained from the light field image data used to synthesize the image can be included in the metadata contained within a JFIF file in accordance with an embodiment of the invention. In many instances, the metadata describing the encoded image can include maps that can be considered to be monochrome images that can be encoded using JPEG encoding. In a number of embodiments, the maps can be compressed using lossless JPEG LS encoding. In several embodiments, the maps can be compressed using lossy JPEG encoding. Utilizing JPEG encoding to compress the maps reduces the size of the maps and enables rendering devices to leverage a JPEG decoder to both decode the image contained within the JFIF file and the maps describing the encoded image. The third byte (i.e. byte 2) of the "DZ Selection Descriptor" indicates the number of sets of metadata describing the encoded image that are contained within the Application marker segment and the fourth byte is reserved. Although specific implementations of the header **2704** describing the metadata contained within the Application marker segment are illustrated in FIGS. **27-30**, any of a variety of implementations can be utilized to identify the maps describing the synthesized image that are present within the metadata contained within an light field image file as appropriate to the requirements of the application in accordance with embodiments of the invention.

Depth Map

Referring back to FIG. **27**, the Application marker segment also includes a "Depth Map Header" **2706** that describes depth map **2716** included within the Application marker segment. The "Depth Map Header" **2706** includes an indication **2708** of the size of "Depth Map Attributes" **2710** included within the "Depth Map Header", the "Depth Map Attributes" **2710**, and a "Depth Map Descriptor" **2712**. As noted above, the depth map **2716** can be considered to be a monochrome image and lossless or lossy JPEG encoding can be utilized to compress the "Depth Map Data" included in a JFIF file.

A "Depth Map Attributes" table in accordance with an embodiment of the invention is illustrated in FIG. **31** and includes information concerning the manner in which the depth map should be used to render the encoded image. In the illustrated embodiment, the information contained within the "Depth Map Attributes" table includes the focal plane and the F# of the synthetic aperture to utilize when rendering the encoded image. Although specific pieces of information related to the manner in which the depth map can be utilized to render the encoded image are illustrated in FIG. **31**, any of a variety of pieces of information appropriate to the requirements of a specific application can be utilized in accordance with embodiments of the invention.

A "Depth Map Descriptor" in accordance with an embodiment of the invention is illustrated in FIG. **32** and includes metadata describing the depth map. In the illustrated embodiment, the "Depth Map Descriptor" includes a zero terminated identifier string "PIDZDH0" and version information. In other embodiments, any of a variety of pieces of information appropriate to the specific requirements of particular applications can be utilized in accordance with embodiments of the invention.

A JFIF Application marker segment is restricted to 65,533 bytes. However, an Application marker can be utilized multiple times within a JFIF file. Therefore, depth maps in accordance with many embodiments of the invention can span multiple APP9 Application marker segments. The manner in which depth map data is stored within an Application marker segment in a JFIF file in accordance with an embodiment of the invention is illustrated in FIG. **33**. In the illustrated embodiment, the depth map data is contained within a descriptor that is uniquely identified using the "PIDZDD0" zero terminated string. The descriptor also includes the length of the descriptor and depth map data.

Although specific implementations of a depth map and header describing a depth map within an Application marker segment of a JFIF file are illustrated in FIGS. **27**, **31**, **32**, and **33**, any of a variety of implementations can be utilized to include a depth map describing an encoded image within a JFIF file as appropriate to the requirements of the application in accordance with embodiments of the invention.

Occlusion Data

Referring back to FIG. **27**, the Application marker segment also includes a "Camera Array Header" **2718** that describes occlusion data **2728** for individual cameras within an array

camera that captured the light field image data utilized to synthesize the image contained within the light field image file. The occlusion data can be useful in a variety of post processing processes including (but not limited) to process that involve modifying the viewpoint of the encoded image. The "Camera Array Header" **2718** includes an indication **2720** of the size of a "Camera Array General Attributes" table **2722** included within the "Camera Array Header", the "Camera Array General Attributes" table **2722**, and a "Camera Array Descriptor" **2724**.

A "Camera Array General Attributes" table in accordance with an embodiment of the invention is illustrated in FIG. **34** and includes information describing the number of cameras and dimensions of a camera array utilized to capture the light field image data utilized to synthesize the image encoded within the JFIF file. In addition, the "Camera Array General Attributes" table can indicate a reference camera position within the array and/or a virtual view position within the array. The "Camera Array General Attributes" table also provides information concerning the number of cameras within the array for which occlusion data is provided within the JFIF file.

A "Camera Array Descriptor" in accordance with an embodiment of the invention is illustrated in FIG. **35** and includes metadata describing the individual camera occlusion data contained within the JFIF file. In the illustrated embodiment, the "Camera Array Descriptor" includes a zero terminated identifier string "PIDZAH0" and version information. In other embodiments, any of a variety of pieces of information appropriate to the specific requirements of particular applications can be utilized in accordance with embodiments of the invention.

In many embodiments, occlusion data is provided on a camera by camera basis. In several embodiments, the occlusion data is included within a JFIF file using an individual camera descriptor and an associated set of occlusion data. An individual camera descriptor that identifies a camera and identifies the number of occluded pixels related to the identified camera described within the JFIF file in accordance with an embodiment of the invention is illustrated in FIG. **36**. In the illustrated embodiment, the descriptor is identified using the "PIDZCD0" zero terminated string. The descriptor also includes a camera number that can be utilized to identify a camera within an array camera that captured light field image data utilized to synthesize the encoded image contained within the JFIF file. In addition, the descriptor includes the number of occluded pixels described in the JFIF file and the length (in bytes) of the data describing the occluded pixels. The manner in which the occluded pixel data can be described in accordance with embodiments of the invention is illustrated in FIG. **36**. The same descriptor "PDIZCD0" is used to identify the occluded pixel data and the descriptor also includes the number of pixels of occluded data contained within the segment, the length of the data in bytes and an offset to the next marker in addition to the occluded pixel data. Due to the restriction on Application marker segments not exceeding 65,533 bytes in data, the additional information enables a rendering device to reconstruct the occluded pixel data across multiple APPS application marker segments within a JFIF file in accordance with embodiments of the invention.

A table describing an occluded pixel that can be inserted within a JFIF file in accordance with an embodiment of the invention is illustrated in FIG. **37**. The table includes the depth of the occluded pixel, the pixel color of the occluded pixel and the pixel coordinates. In the illustrated embodiment, the pixel color is illustrated as being in the RGB domain. In other embodiments, the pixel color can be expressed in any domain including the YCbCr domain.

Although specific implementations for storing information describing occluded pixel depth within an Application marker segment of a JFIF file are illustrated in FIGS. **27**, **34**, **35**, and **36**, any of a variety of implementations can be utilized to include occluded pixel information within a JFIF file as appropriate to the requirements of the application in accordance with embodiments of the invention.

Auxiliary Maps

Referring back to FIG. **27**, any of a variety of auxiliary maps can be included in an Application marker segment within a JFIF file in accordance with an embodiment of the invention. The total number of auxiliary maps and the types of auxiliary maps can be indicated in the Application marker segment. Each auxiliary map can be expressed using an "Auxiliary Map Descriptor" **2732** and "Auxiliary Map Data" **2734**. In the illustrated embodiment, the "Auxiliary Map Descriptor" **2732** is included in an "Auxiliary Map Header" **2730** within the Application marker segment in the JFIF file.

An "Auxiliary Map Descriptor" that describes an auxiliary map contained within a light field image file in accordance with an embodiment of the invention is illustrated in FIG. **39**. The "Auxiliary Map Descriptor" includes an identifier, which is the "PIDZAM0" zero terminated string and information specifying the type of auxiliary map and number of bits per pixel in the map. As noted above, any of a variety of auxiliary maps derived from light field image data used to synthesize an encoded image can be included within a JFIF file in accordance with embodiments of the invention. In the illustrated embodiment, confidence maps, silhouette edge maps, regular edge maps, and missing pixel maps are supported.

"Auxiliary Map Data" stored in a JFIF file in accordance with an embodiment of the invention is conceptually illustrated in FIG. **40**. The "Auxiliary Map Data" uses the same "PDIZAD0" zero terminated string identifier and includes the number of pixels of the auxiliary map contained within the segment, the length of the data in bytes and an offset to the next marker in addition to pixels of the auxiliary map. Due to the restriction on Application marker segments not exceeding 65,533 bytes in data, the additional information enables a rendering device to reconstruct the auxiliary map describing the encoded image across multiple APPS application marker segments within a JFIF file.

Although specific implementations for storing auxiliary maps within an Application marker segment of a JFIF file are illustrated in FIGS. **27**, **39**, and **40**, any of a variety of implementations can be utilized to include auxiliary map information within a JFIF file as appropriate to the requirements of the application in accordance with embodiments of the invention. Various examples of auxiliary maps that can be utilized to provide additional information concerning an encoded image based upon the light field image data utilized to synthesize the encoded image in accordance with embodiments of the invention are discussed below.

Confidence Maps

A confidence map can be utilized to provide information concerning the relative reliability of the information at a specific pixel location. In several embodiments, a confidence map is represented as a complimentary one bit per pixel map representing pixels within the encoded image that were visible in only a subset of the images used to synthesize the encoded image. In other embodiments, a confidence map can utilize additional bits of information to express confidence using any of a variety of metrics including (but not limited to)

a confidence measure determined during super resolution processing, or the number of images in which the pixel is visible.

Edge Maps

A variety of edge maps can be provided included (but not limited to) a regular edge map and a silhouette map. A regular edge map is a map that identifies pixels that are on an edge in the image, where the edge is an intensity discontinuity. A silhouette edge maps is a map that identifies pixels that are on an edge, where the edge involves an intensity discontinuity and a depth discontinuity. In several embodiments, each can be expressed as a separate one bit map or the two maps can be combined as a map including two pixels per map. The bits simply signal the presence of a particular type of edge at a specific location to post processing processes that apply filters including (but not limited to) various edge preserving and/or edge sharpening filters.

Missing Pixel Maps

A missing pixel map indicates pixel locations in a synthesized image that do not include a pixel from the light field image data, but instead include an interpolated pixel value. In several embodiments, a missing pixel map can be represented using a complimentary one bit per pixel map. The missing pixel map enables selection of post-processing filters to improve image quality. In many embodiments, a simple interpolation algorithm can be used during the synthesis of a higher resolution from light field image data and the missing pixels map can be utilized to apply a more computationally expensive interpolation process as a post processing process. In other embodiments, missing pixel maps can be utilized in any of a variety of different post processing process as appropriate to the requirements of a specific application in accordance with embodiments of the invention.

Processes for Rendering Images Using Light Field Image Files

As noted above, rendering a light field image file can be as simple as decoding an encoded image contained within the light field image file or can involve more complex post processing of the encoded image using metadata derived from the same light field image data used to synthesize the encoded image. A process for rendering a light field image in accordance with an embodiment of the invention is illustrated in FIG. **41**. The process **4100** includes parsing (**4102**) the light field image file to locate the encoded image contained within the image file. The encoded image file is decoded (**4104**). As noted above, the image can be encoded using a standards based encoder and so the decoding process can utilize a standards based codec within a rendering device, or the image can be encoded using a proprietary encoding and a proprietary decoder is provided on the rendering device to decode the image. When the process for rendering the image simply involves rendering the image, the decoded image can be displayed. When the process for rendering the image includes post processing, the image file is parsed (**4106**) to locate metadata within the file that can be utilized to perform the post processing. The metadata is decoded (**4108**). The metadata can often take the form of maps that can be encoded using standards based image encoders and a standards based decoder present on the rendering device can be utilized to decode the metadata. In other embodiments, a proprietary decoding process is utilized to decode the metadata. The metadata can then be used to perform (**4110**) the post processing of the encoded image and the image can be displayed (**4112**). The display of the image can be local. Alternatively the image can be streamed to a remote device or encoded as an image and provided to a remote device for display.

Although specific processes for rendering an image from a light field image file are discussed with reference to FIG. **41**, any of a variety of processes appropriate to the requirements of a specific application can be utilized to render an image for display using a light field image file in accordance with an embodiment of the invention. As noted above, any of a variety of standards based encoders and decoders can be utilized in the encoding and decoding of light field image files in accordance with embodiments of the invention. Processes for rendering images using light field image files that conform to the JFIF standard and include an image and/or metadata encoded in accordance with the JPEG standard are discussed further below.

Processes for Rendering Images from JFIF Light Field Image Files

Processes for rending images using light field image files that conform to the JFIF standard can utilize markers within the light field image file to identify encoded images and metadata. Headers within the metadata provide information concerning the metadata present in the file and can provide offset information or pointers to the location of additional metadata and/or markers within the file to assist with parsing the file. Once appropriate information is located, a standard JPEG decoder implementation can be utilized to decode encoded images and/or maps within the file.

A process for displaying an image rendered using a light field image file that conforms to the JFIF standard using a JPEG decoder in accordance with an embodiment of the invention is illustrated in FIG. **42**. The process **4200** involves parsing (**4202**) the light field image file to locate a Start of Image (SOI) Marker. The SOI marker is used to locate an image file encoded in accordance with the JPEG format. The encoded image can be decoded (**4204**) using a JPEG decoder. When no post processing of the decoded image is desired, the image can simply be displayed. Where post processing of the image is desired (e.g. to change the view point of the image and/or the focal plane of the image), the process parses (**4206**) the light field image file to locate an appropriate Application marker. In the illustrated embodiment, an APP9 marker indicates the presence of metadata within the light field image file. The specific metadata within the file can be determined by parsing (**4206**) a header within the APP9 Application marker segment that describes the metadata within the file. In the illustrated embodiment, the header is the "DZ Header" within the APP9 Application marker segment. The information within the metadata header can be utilized to locate (**4208**) specific metadata utilized in a post processing process within the light field image file. In instances where the metadata is encoded, the metadata can be decoded. In many embodiments, metadata describing an encoded image within a light field image file is in the form of a map that provides information concerning specific pixels within an encoded image contained within the light field image file and JPEG encoding is used to compress the map. Accordingly, a JPEG decoder can be utilized to decode the map. The decoded metadata can be utilized to perform (**4212**) a post processes the decoded image. The image can then be displayed (**4214**). In many embodiments, the image is displayed on a local display. In a number of embodiments, the image is streamed to a remote display or encoded as an image and forwarded to a remote device for display.

Although specific processes for displaying images rendered using light field image files are discussed above with respect to FIG. **42**, any of a variety of processes for parsing a light field image file and decoding images and/or metadata encoded in accordance with the JPEG standard using a JPEG decoder can be utilized in accordance with embodiments of

the invention. Much of the discussion above references the use of metadata derived from light field image data and contained within a light field image file to perform post processing processes on an encoded image synthesized from the light field image data.

Reducing Computational Complexity

A variety of strategies can be utilized to reduce the computational complexity of the processes outlined above for determining depth maps and for determining the visibility of images captured by a camera array. In several embodiments, a depth map is constructed by only searching for depth at a reduced (i.e. sparse) subset of pixel locations. The depth search is done at fewer points (i.e. a sparser set of points in the image) and for points that depth is not calculated, the depth is assigned through other means. By the end, this sparse depth search provides a depth for every pixel location in a reference image where some pixels are searched and others are filled in through interpolation. As previously stated, not every pixel in the final depth map has a depth obtained by comparing the similarity of the pixel to corresponding pixels in the captured images. Instead, in regions where no correspondence search is done, the depths of many of the pixels are determined using processes including (but not limited to) averaging the depths of surrounding pixels (where the correspondence search has been run) or interpolating the depths of adjacent pixels which have been calculated. By reducing the number of pixels for which depth measurements are performed, the amount of computation used to generate a depth map can be reduced. In several embodiments, the amount of computation used when detecting a depth map can also be reduced by detecting textureless areas of the image and using processes including (but not limited to) assigning a single depth value from the nearest indicator pixel where depth has been calculated, averaging the depths of surrounding pixels or interpolating the depths of adjacent pixels to determine the depth of pixels in the textureless areas of the image. In other embodiments, any of a variety of processes for reducing the computational complexity of generating a depth map can be utilized as appropriate to the requirements of specific applications in accordance with embodiments of the invention including varying the precision of the depth estimates within the depth map based upon characteristics of the scene including (but not limited to) regions containing edges, and/or based upon object distance. Processes for generating depth maps from sparse depth searches and for detecting textureless regions in images in accordance with embodiments of the invention are discussed further below.

Generating Depth Maps from Sparse Depth Search

Processes for generating depth maps through sparse search in accordance with embodiments of the invention typically involve determining depth of a sparse set of pixels spaced or distributed throughout the reference image. Based upon this initial depth map consisting of sparse points, depth transitions can be detected and the depths of pixels surrounding the depth transitions can be directly measured using the processes outlined above. The depths of the remaining pixels can be determined based upon the depths of sparse pixels in the depth map. In many embodiments, the depth measurements are performed using a subset of the pixels in the reference image at the resolution at which they were captured.

A process for determining a depth map through sparse search in accordance with an embodiment of the invention is illustrated in FIG. **13**. The process **1300** includes dividing (**1302**) the reference image into spatial blocks (or groups of associated pixels) and generating (**1304**) depth measurements for a sparser subset of indicator pixels within the spatial blocks. Here, spatial block may be taken to refer interchangeably to a rectangular block of pixels, or a subset of associated pixels that need not conform to any particular shape.

Indicator pixels are a subset of the pixels within the spatial block (or group of associated pixels) and are typically selected to provide information concerning variation in depth across the spatial block. A spatial block **1400** including a plurality of indicator pixels **1402** in accordance with an embodiment of the invention is illustrated in FIG. **14**. The indicator pixels **1402** are selected at the edges and at the center of the spatial block. Although specific indicator pixels are illustrated in FIG. **14**, the arrangement of indicators within a spatial block or group of associated pixels can be varied and/or any of a variety of pixels within a spatial block can be selected as indicator pixels as appropriate to the requirements of a specific application. In a number of embodiments, different shaped spatial blocks are utilized and the shape of the spatial block can be varied. In several embodiments, the arrangement of indicator pixels within the spatial blocks can be varied. In many embodiments, the indicator pixels are selected based upon scene content. In certain embodiments, the indicator pixels are selected based on which points within the spatial block have the highest SNR in the reference image to increase the likelihood that the points most likely to give confident depth results are used. In another embodiment, fixed spatial positions are chosen for some indicator pixels (as indicated in FIG. **14**) for all blocks, and some subset of indicator pixels are assigned to points with highest SNR in the spatial block (i.e. a mixed configuration). In another embodiment, a segmentation process can be used to create relevant spatial regions based on scene content. Although a rectangular spatial block is shown other techniques could be used for splitting the image into spatial clusters, which contain some indicator pixels as described above. Furthermore spatial blocks can be larger in certain portions of the image than in others.

Referring back to FIG. **13**, depth can be assigned (**1306**) to the pixels in each block based upon the depths of the indicator pixels. In several embodiments, the assigned depth is obtained through interpolation of the neighboring indicator pixels. In several embodiments, the depth of a non-indicator pixel may be calculated as a normalized weighted average of the distances to the nearest indicator pixels within a fixed neighborhood. Alternatively, nearest neighbor interpolation (**1308**) can be utilized to assign depths to the pixels in the spatial block based upon the depth measurements of the indicator pixels. In another embodiment, weights for the interpolation can incorporate intensity similarity as well as spatial distance to the nearest indicator pixels. In another embodiment, a non-linear regression such as (but not limited to) a Kernel Regression may be used to fill in the missing positions between depths sampled at the indicator pixel positions. In another embodiment, a single depth for the entire block is assigned by minimizing the summed costs of the indicator pixels within the block. In other embodiments, any of a variety of techniques can be utilized to generate depth information for pixels within a spatial block.

In many embodiments, the reliability of each of the spatial blocks in the depth map is determined (**1310**). Within the spatial block, depths will have been estimated both for indicator pixels (where search has occurred) and non-indicator pixels (where depths have been interpolated based on indicator pixel results). For the indicator and non-indicator pixels, the costs of the estimated depths within the block are determined. The costs of each pixel in the block are summed to create a reliability indicator. If the total cost of all pixels within the block is greater than a threshold, then the spatial block is marked as unreliable due to the fact that the estimated

depths for some pixels appear to have poor correspondence. Where a spatial block has been determined to have low reliability of poor spatial correspondence, then the block is likely to contain a depth transition or occlusion. If such is the case, then the full correspondence search and occlusion processing can be run within the spatial block.

If a spatial block is determined to have a depth transition per the criteria above, then the spatial block may be 'split' and new sets indicator pixels selected in each of the two child spatial blocks and the process iterated. In one embodiment, the block may be split in half. In another embodiment, the block may be split into unequal regions depending on the depths solved by the indicator pixels within the spatial block.

Where depth transitions are detected within and/or between spatial blocks, the depth map can be refined (**1312**) by performing additional depth measurements within the spatial block that contains the depth transitions. In this way, the computational complexity of generating the depth map is reduced by reducing the number of depth measurements performed in generating an accurate depth map.

Although a specific process for generating a depth map from sparse searches in accordance with embodiments of the invention is illustrated in FIG. **13**, any of a variety of processes that generate a depth map by performing fewer depth measurements in regions of similar or slowly transitioning depth can be utilized in accordance with embodiments of the invention.

Reducing Computation in Textureless Regions of Images

In many embodiments, the process of generating a depth map involves reducing the amount of computation needed for textureless regions of the image. Textureless areas can be ambiguous with parallax, because the corresponding pixels at many hypothesized depths may be similar. Therefore, depth measurements within a textureless area can generate unreliable and noisy results. In many embodiments, the SNR in the region surrounding a pixel is used when determining the depth of the pixel to identify whether the pixel is in a textureless area. An initial depth estimate or a set of initial depth estimates for a given pixel can be determined based upon the depth of at least one adjacent pixel for which a depth has previously been determined. When the variance of the corresponding pixels for the given pixel (or any other similarity measure) is below the SNR threshold in the region surrounding the pixel, the pixel can be assumed to be part of a textureless area and (one of) the approaches described below can be used to select the depth of pixel. Otherwise, a depth measurement can be performed using a process similar to the processes described above.

In many embodiments, textureless regions may be detected using a fixed threshold on the SNR. The computation for the search in such regions may be reduced by reducing the number of depths searched. In many embodiments, the full set of depths will be searched until a minimum cost depth is identified that is below a noise-dependent threshold that takes into account the noise characteristics of the sensor. When the minimum cost is found to be below the threshold the depth is accepted as the depth of the textureless region and no more depths are searched (i.e. the search is terminated as soon as a depth that has "close enough" correspondence is found). In many embodiments, the search in textureless regions may save computation by searching the full range of disparity but at larger increments than are done in the normal search for a region with texture (i.e. reducing the number of depths searched)—the best cost will be selected as the depth of the pixel in the textureless region.

A process for detecting textureless regions using the SNR surrounding a pixel in accordance with an embodiment of the invention is illustrated in FIG. **15**. The process **1500** includes selecting (**1502**) a pixel from the reference image and detecting (**1504**) the SNR in the region around the selected pixel. An initial hypothesized depth d can be determined (**1506**) for the pixel. In many embodiments, the initial hypothesized depth d is determined based upon the depth of one or more pixels in the region surrounding the selected pixel. A determination (**1508**) is then made concerning whether the variance or cost of the corresponding pixels at the hypothesized depth is below a threshold that can be (but is not limited to) predetermined or a function of the SNR in the region surrounding the selected pixel. In other embodiments, any of a variety of similarity measures can be utilized to determine whether the region surrounding the pixel is textureless. In the event that variance or cost of the corresponding pixels is below a noise or predetermined threshold, then the hypothesized depth is selected as the most likely depth on the assumption that the pixel is located within a textureless region. When the variance or cost of the corresponding pixels exceeds the noise or predetermined threshold, then the depth of a pixel is determined in accordance with a process similar to the processes described above.

Although a specific process for detecting textureless areas within a reference image are described above with respect to FIG. **15**, any of a variety of processes for detecting textureless areas in an image can be utilized in accordance with embodiments. Furthermore, any of a variety of processes can be utilized to detect other characteristics of an image that can be relied upon to reduce the number of depth measurements that are made in generating a reliable depth map in accordance with embodiments of the invention.

Generating Depth Maps from Virtual Viewpoints

While much of the discussion provided above describes the generation of depth maps with respect to images captured by a reference camera, systems and methods in accordance with embodiments of the invention can synthesize images from virtual viewpoints. A virtual viewpoint is a reference viewpoint that does not correspond to the viewpoint of any cameras within a camera array. Accordingly, irrespective of the number of color channels within a camera array, none of the color channels include a camera in which image data is captured from the reference viewpoint. An example of a virtual viewpoint that can be defined in accordance with an embodiment of the invention is illustrated in FIG. **12**. The array camera module **1200** includes a 4×4 array of cameras including 8 Green cameras, 4 Red cameras, and 4 Blue cameras. A virtual camera **1202** is defined with a virtual viewpoint at the center of the array camera module. Although a specific virtual viewpoint is illustrated in FIG. **12**, any virtual viewpoint can be arbitrarily defined with respect to the cameras within a camera array.

When determining a virtual viewpoint depth map, there is no explicit reference camera which can be searched and used for cost metric comparisons. In many embodiments, the depth of a given pixel (x, y) in an image synthesized from the virtual viewpoint is determined by calculating the effective baseline from the virtual imager with respect to all other cameras in the array. The baseline for a camera at position (i,j) with respect to the virtual viewpoint would be $(i,j)-(i_v,j_v)$ where (i_v,j_v) is the location of the virtual viewpoint **1202**. Once the baselines between the individual cameras is determined with respect to the virtual viewpoint, the process of estimating depth proceeds by searching for depths at which corresponding pixels having the highest similarity. For each pixel (x, y) in the virtual reference camera (i.e. an image from the virtual viewpoint), the search can proceed much as in the typical parallax scenario, where for each depth d, the disparity with respect to

each of the alternate view cameras is determined at that depth, and then the similarity of corresponding pixels in one or more of the color channels is determined using an appropriate cost metric. In many embodiments, the combination cost metric described above for determining the similarity of pixels in color channels that do not contain the reference camera can be utilized. In many embodiments, a camera adjacent the virtual viewpoint in a specific color channel can be used as a reference camera for the purpose of determining the similarity of the pixel in the chosen reference camera with corresponding pixels in image data captured by other cameras in the color channel. In many embodiments, a Green camera is chosen as a reference camera for the purpose of determining the similarity of corresponding Green pixels and a combination cost metric is used for corresponding pixels in other color channels. In many embodiments, the process of determining an initial depth map for the virtual viewpoint can involve forming groups of cameras corresponding to patterns of visibility within the scene in a similar manner to that described above with respect to FIGS. **8**A-**8**I.

A depth map generated for a virtual viewpoint can be utilized to synthesize a high resolution image from a virtual viewpoint using a super-resolution process in accordance with embodiments of the invention. The primary difference in the synthesis of a high resolution image from a virtual viewpoint is that the high resolution grid is from a virtual viewpoint, and the pixels are fused to the high resolution grid using correspondences calculated with baselines which are with respect to the virtual view position and not a physical reference camera. In this case there is no physical reference camera having pixels that map regularly on the high resolution grid. As can be readily appreciated, processes for determining confidence maps for depth maps with respect to virtual viewpoints can be determined using similar accommodations related to analyzing the synthesized reference image or choosing an image close to the virtual viewpoint as a proxy for performing SNR and/or other related measurements. Although specific processes for generating depth maps with respect to virtual viewpoints are described above, any of a variety of processes incorporating the cost metrics and techniques outlined above can be utilized to generate depth estimates for virtual viewpoints in accordance with embodiments of the invention. Systems for performing parallax detection and correction and for generating depth maps in accordance with embodiments of the invention are discussed further below.

Systems for Performing Parallax Detection

A system for generating a depth map and visibility information using processes similar to those described above is illustrated in FIG. **16**. The system includes a parallax detection module **1600** that takes as an input captured images that form a light field and calibration information for an array camera and outputs a depth map, and the estimated visibility of the pixels of the captured images. In many embodiments, the parallax detection module **1600** also outputs a confidence map indicating the reliability of the depth measurements for specific pixels within the reference image. As is discussed further below, the depth map, estimated visibility information, and/or confidence map can be provided to a super-resolution processing module within an array camera to generate a higher resolution image from the captured images and to any of a variety of applications that can utilize depth, confidence and/or visibility information. In many embodiments, the parallax detection module and the super-resolution module are implemented in software and/or firmware on a microprocessor within the array camera. In several embodiments, the software associated with the parallax detection module and the super-resolution module is stored within memory within the array camera. In other embodiments, the parallax detection module and/or the super-resolution module can be implemented using any appropriately configured hardware and/or software. The generation of high resolution images from a light field captured by an array camera using a depth map generated in accordance with embodiments of the invention is discussed further below.

Super-Resolution Processing Using Depth Maps

As is noted in U.S. patent application Ser. No. 12/967,807 (incorporated by reference above) disparity between images can introduce significant artifacts when performing super-resolution processing. Therefore, the super-resolution processes disclosed in U.S. patent application Ser. No. 12/967,807 can involve applying scene dependent geometric corrections to the location of each of the pixels in the images captured by an array camera prior to using the images to synthesize a higher resolution image. The baseline and back focal length of the cameras in an array camera can be readily determined, therefore, the unknown quantity in estimating the scene dependent geometric shifts observed in the captured images is the distance between the array camera and different portions of the scene. When a depth map and visibility information is generated in accordance with the processes outlined above, the scene dependent geometric shifts resulting from the depths of each of the pixels can be determined and occluded pixels can be ignored when performing super-resolution processing. In many embodiments, a confidence map is generated as part of the process of generating a depth map and the confidence map is provided as an input to the super-resolution process to assist the super-resolution process in evaluating the reliability of depth estimates contained within the depth map when performing fusion of the pixels from the input images.

A process for generating a high resolution image using a light field captured by an array camera involving the generation of a depth map in accordance with an embodiment of the invention is illustrated in FIG. **17**. The process **1700** involves capturing (**1702**) a light field using an array camera and selecting (**1704**) a reference viewpoint that can be utilized to synthesize a high resolution image. In many embodiments, the reference viewpoint is predetermined based upon the configuration of the array camera. In a number of embodiments, calibration information is utilized to increase (**1706**) the correspondence between captured images. In many embodiments, the correspondence between captured images involves resampling the images. An initial depth map is determined (**1708**) and occlusions are determined and used to update (**1710**) the depth map. In several embodiments, the process of detecting occlusions and updating the depth map is iterative.

In a number of embodiments, the depth map is utilized to generate (**1712**) information concerning the visibility of the pixels within the captured light field from the reference viewpoint. In several embodiments, a confidence map is (optionally) generated (**1713**) with respect to the depth estimates contained within the depth map and the depth map, the visibility information, and/or the confidence map are provided (**1714**) to a super-resolution processing pipeline. In several embodiments, the super-resolution processing pipeline is similar to any of the super-resolution processing pipelines disclosed in U.S. patent application Ser. No. 12/967,807. The super-resolution processing pipeline utilizes information including the light field, the depth map, the visibility information, and the confidence map to synthesize (**1718**) a high resolution image from the reference viewpoint, which is output (**1718**) by the array camera. In several embodiments, the

process of synthesizing a higher resolution image involves a pilot fusion of the image data from the light field onto a higher resolution grid. The result of the pilot fusion can then be utilized as a starting point to synthesize a higher resolution image using the super-resolution process.

As is described in U.S. patent application Ser. No. 12/967,807, the process illustrated in FIG. **17** can be performed to synthesize a stereoscopic 3D image pair from the captured light field. Although a specific process for synthesizing a high resolution image from a captured light field is illustrated in FIG. **17**, any of a variety of processes for synthesizing high resolution images from captured light fields involving the measurement of the depth of pixels within the light field can be utilized in accordance with embodiments of the invention.

While the above description contains many specific embodiments of the invention, these should not be construed as limitations on the scope of the invention, but rather as an example of one embodiment thereof. Accordingly, the scope of the invention should be determined not by the embodiments illustrated, but by the appended claims and their equivalents.

What is claimed is:

1. A camera array, comprising:
- a plurality of cameras configured to capture images of a scene from different viewpoints;
- a processor; and
- memory containing an image processing application;
- wherein the image processing application stored in memory directs the processor to:
  - separately configure the imaging parameters for each of the plurality of cameras;
  - read out image data from the plurality of cameras including a set of images captured from different viewpoints;
  - store the image data in the memory;
  - select a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints;
  - normalize the set of images to increase the similarity of corresponding pixels within the set of images;
  - determine initial depth estimates for pixel locations in an image from the reference viewpoint based upon the disparity at which corresponding pixels in the set of images have the highest degree of similarity;
- compare the similarity of the corresponding pixels in the set of images to detect mismatched pixels;
- when an initial depth estimate does not result in the detection of a mismatch between corresponding pixels in the set of images, selecting the initial depth estimate as the depth estimate for the pixel location in the image from the reference viewpoint; and
- when an initial depth estimate results in the detection of a mismatch between corresponding pixels in the set of images, updating the depth estimate for the pixel location in the image from the reference viewpoint by:
  - determining a set of candidate depth estimates using a plurality of competing subsets of the set of images based upon the disparities at which corresponding pixels in each of a plurality of competing subsets of images have the highest degree of similarity; and
  - selecting the candidate depth of the subset having the corresponding pixels with the highest degree of similarity as the updated depth estimate for the pixel location in the image from the reference viewpoint.

2. The camera array of claim **1**, wherein the optics within each camera are configured so that the pixels of the camera sample the same object space with sub-pixel offsets.

3. The camera array of claim **2**, wherein the image processing application further directs the processor to:
- determine the visibility of the pixels in the set of images from the reference viewpoint by:
  - identifying corresponding pixels in the set of images using the current depth estimates; and
  - determining that a pixel in a given image is not visible in the reference viewpoint when the pixel fails a photometric similarity criterion determined based upon a comparison of corresponding pixels; and
- fuse pixels from the set of images using the depth estimates to create a fused image having a resolution that is greater than the resolutions of the images in the set of images by:
  - identifying the pixels from the set of images that are visible in an image from the reference viewpoint using the visibility information;
  - applying scene dependent geometric shifts to the pixels from the set of images that are visible in an image from the reference viewpoint to shift the pixels into the reference viewpoint, where the scene dependent geometric shifts are determined using the depth estimates; and
  - fusing the shifted pixels from the set of images to create a fused image from the reference viewpoint having a resolution that is greater than the resolutions of the images in the set of images.

4. The camera array of claim **3**, wherein the image processing application further directs the processor to synthesize an image from the reference viewpoint by performing a super-resolution process based upon the fused image from the reference viewpoint, the set of images captured from different viewpoints, the depth estimates, and the visibility information.

5. The camera array of claim **1**, wherein at least one camera in the plurality of cameras is a monochrome camera.

6. The camera array of claim **5**, wherein at least one camera includes a Bayer color filter pattern.

7. The camera array of claim **5**, wherein the image processing application further directs the processor to select the viewpoint of a camera in the plurality of cameras that captures image data in multiple color channels as the reference viewpoint.

8. The camera array of claim **1**, wherein at least one camera in the plurality of cameras captures image data in multiple color channels.

9. The camera array of claim **1**, wherein different cameras within the plurality of cameras capture image data with respect to different portions of the electromagnetic spectrum.

10. The camera array of claim **1**, wherein different cameras within the array camera module capture image data with respect to different portions of the electromagnetic spectrum.

11. The camera array of claim **1**, wherein the array camera module further comprises light filters within each optical channel formed by the lens stacks.

12. The camera array of claim **11**, wherein the light filters are applied to the pixels of the focal planes.

13. A camera array, comprising:
- an array camera module comprising:
  - an imager array including an array of focal planes, where each focal plane includes a plurality of rows of pixels that also forms a plurality of columns of pixels, and each focal plane is contained within a region of the imager that does not contain pixels from another focal plane; and
  - an optic array including an array of lens stacks, where each lens stack creates an optical channel that forms

an image of the scene on an array of pixels within a corresponding focal plane;

a processor; and

memory containing an image processing application;

wherein the image processing application stored in memory directs the processor to:

independently control the imaging parameters of the focal planes in the array camera module;

read out image data from the array camera module forming a set of images captured from different viewpoints;

store the image data in the memory;

select a reference viewpoint relative to the viewpoints of the set of images captured from different viewpoints;

normalize the set of images to increase the similarity of corresponding pixels within the set of images;

determine initial depth estimates for pixel locations in an image from the reference viewpoint based upon the disparity at which corresponding pixels in the set of images have the highest degree of similarity;

compare the similarity of the corresponding pixels in the set of images to detect mismatched pixels;

when an initial depth estimate does not result in the detection of a mismatch between corresponding pixels in the set of images, selecting the initial depth estimate as the depth estimate for the pixel location in the image from the reference viewpoint; and

when an initial depth estimate results in the detection of a mismatch between corresponding pixels in the set of images, updating the depth estimate for the pixel location in the image from the reference viewpoint by:

determining a set of candidate depth estimates using a plurality of competing subsets of the set of images based upon the disparities at which corresponding pixels in each of a plurality of competing subsets of images have the highest degree of similarity; and

selecting the candidate depth of the subset having the corresponding pixels with the highest degree of similarity as the updated depth estimate for the pixel location in the image from the reference viewpoint.

14. The camera array of claim **13**, wherein the lens stack within each optical channel is configured so that the pixels of each focal plane sample the same object space with sub-pixel offsets.

15. The camera array of claim **14**, wherein the image processing application further directs the processor to:

determine the visibility of the pixels in the set of images from the reference viewpoint by:

identifying corresponding pixels in the set of images using the current depth estimates; and

determining that a pixel in a given image is not visible in the reference viewpoint when the pixel fails a photometric similarity criterion determined based upon a comparison of corresponding pixels; and

fuse pixels from the set of images using the depth estimates to create a fused image having a resolution that is greater than the resolutions of the images in the set of images by:

identifying the pixels from the set of images that are visible in an image from the reference viewpoint using the visibility information; and

applying scene dependent geometric shifts to the pixels from the set of images that are visible in an image from the reference viewpoint to shift the pixels into the reference viewpoint, where the scene dependent geometric shifts are determined using the depth estimates; and

fusing the shifted pixels from the set of images to create a fused image from the reference viewpoint having a resolution that is greater than the resolutions of the images in the set of images.

16. The camera array of claim **15**, wherein the image processing application further directs the processor to synthesize an image from the reference viewpoint by performing a super-resolution process based upon the fused image from the reference viewpoint, the set of images captured from different viewpoints, the depth estimates, and the visibility information.

17. The camera array of claim **13**, wherein at least one camera in the array camera module is a monochrome camera.

18. The camera array of claim **17**, wherein at least one camera in the array camera module includes a Bayer color filter pattern.

19. The camera array of claim **17**, wherein the image processing application further directs the processor to select the viewpoint of a camera in the plurality of cameras that captures image data in multiple color channels as the reference viewpoint.

20. The camera array of claim **13**, wherein at least one camera in the array camera module captures image data in multiple color channels.

* * * * *